

Dear Mr Mbeki . . .

An open letter to the president of South Africa.

We are writing in response to your recent letter to world leaders, including US president Bill Clinton and United Nations secretary general Kofi Annan, expressing your concern about the horrific situation that your country faces over the spread of AIDS, and your desire to see the situation approached in the most rigorously scientific way possible (see page 611). We share this goal; AIDS will not be defeated or contained without access to the best treatment that modern science has to offer. But we are also concerned that, in your admirable enthusiasm to ensure that a wide spectrum of scientific views is heard, you appear tempted to give greater weight to some voices than the scientific process justifies.

No one who has been impressed by your success in what many claimed impossible — the peaceful transition to democracy in South Africa following the long struggle against the iniquities of apartheid — will reject the argument that there are times when the voice of those challenging the existing order must be heard. Your own colleagues have referred to the astronomer Galileo in this context. But this does not mean that all dissidents and 'heretics' can claim equal legitimacy merely on the basis of their persecution; democratically endorsed procedures exist through which their ideas can be put to the test, and viable heresies separated from those that, after close scrutiny, deserve to be placed aside.

Politics has developed one set of such procedures: the ballot box, parliamentary debate and constitutional law. Science has developed its own, very different, set. Contrary to the impression given by your letter, science thrives on the ideas of heretics. But heretical hypotheses only become widely accepted in science if they prove useful and effective in understanding and interacting with the natural world. The peer-review system is little more than a way of speeding up the process

of sorting out those ideas which have a greater chance than others of surviving intellectual scrutiny and testing through experiment.

One hypothesis that has survived this process is that the idea that there is a direct, causal relationship between infection with the human immunodeficiency virus (HIV) and the onset of AIDS. We are well aware of the arguments of those who challenge such a direct relationship. Our columns have been — and remain — open to anyone offering evidence to the contrary, but on one simple condition: that their evidence passes the same rigorous tests of scientific robustness that are applied to any scientific paper that we receive. So far this has not happened. Those who have experienced rejection may choose to castigate this as 'censorship', but the vast majority of authors of the scientific papers that we reject on technical grounds accept the process as valid and necessary for the health of science.

You yourself admit in your letter to President Clinton and the other world leaders that your comments about the treatment being given to heretical ideas on the nature of AIDS may be "extravagant"; you justify this on the grounds that in the recent past you have had, in your own words, "to fix our eyes on the very face of tyranny". But as Koïchiro Matsuura, the new head of Unesco, said at a meeting in Nigeria this week: "Without a scientific capacity of its own, Africa will not be able to tackle and overcome its endemic diseases".

Mr Mbeki, we ask you, in the spirit of Matsuura's comment, not to ignore the advice of your own leading scientific and medical experts, nor to reject those aspects of science that offer your country the greatest hope for the future. A respect for vigorous scientific debate is one of these aspects that we endorse. Giving excessive credence to populist hypotheses that fly in the face of established evidence and fail to survive rigorous peer review is not. ■

Big can still be great

High-energy physicists in the United States need to take time to reach agreement on new goals.

The profound challenges of high-energy physics continue to attract some of the best minds. Its experimental successes have been triumphs of human ingenuity, its costs by and large tightly controlled, its discoveries a wonderful blend of the predicted and the wholly unexpected. Now it is on the threshold of a vast new landscape whose inhabitants, putative 'supersymmetric' partners of the known elementary particles, and their relationships can only dimly be perceived. Recent reported detections of oscillations between components of neutrinos and the imminent exploration of the predicted Higgs particles point to new phenomena to be explored over the next few years. A decade or more from now, the next lepton collider will be needed to define that supersymmetric landscape with the required precision. Unforeseen discoveries can also be expected.

Such progress costs substantial but fully defensible sums of money. But in the United States, the political climate is not propitious for expensive international science. Furthermore, it is proving difficult (see page 909) for the US high-energy physics community to achieve consensus about the best path forward. That is not surprising

given the difficult technical trade-offs between different approaches, the disruption and demoralization that followed Congress's 1993 decision to abandon the Superconducting Super Collider, and the current budget squeeze on the discipline. The last factor is creating an environment of fear, in which researchers and laboratories may put a short-term desire to protect their projects ahead of the greater long-term requirement to rally around an agreed future programme.

Such a programme is required if the community is to make its case effectively. The perception that high-energy physics will be stuck in a holding pattern until the completion of the European Large Hadron Collider (LHC) has taken hold among some influential officials in Washington and, should it gain currency, it will result in more frozen budgets, staff cuts and division of the community. It is understandable that the Department of Energy has asked for a clearer consensus. But this should not be forced. A gathering of the community at Snowmass, Colorado, in summer 2001 is probably the best opportunity for it to coalesce. Congress and the administration should ensure that the current programme is adequately funded at least until then. ■





Screen test
Zebrafish hold a key
to developmental
genetics
p910



AIDS awareness
Mbeki faces new
storm over HIV
remarks
p911



Gene fragments
The Human Genetic
Diversity Project
struggles to survive
p912



Terra firma
Earth observation
satellite put through
its paces
p913

Budget crisis forces hard choices on US high-energy physics

Washington

US high-energy physicists, facing deep budget cuts and divided on their priorities, have been given six months to come up with a consensus on the future direction of their \$700 million research programme.

Peter Rosen, head of the high energy and nuclear physics office at the Department of Energy, asked the department's High Energy Physics Advisory Panel (HEPAP) last month to revisit a report that it produced in 1998 on the future direction of the discipline in the United States.

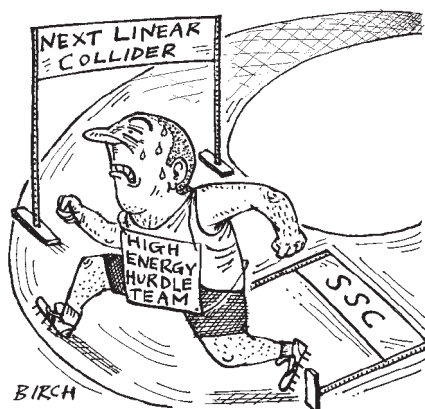
Rosen believes the new study is needed by this autumn to shore up support for the high-energy physics programme in Washington. Officials in both Congress and the Clinton administration have expressed concern about the what they perceive as a lack of clarity about the programme's goals.

"There's a lot of questions about whether the community really deserves the funding which it asks for," says Michael Lubell, head of public affairs at the American Physical Society in Washington. Lubell says that the field badly needs to make more of the spin-offs from particle physics — such as the invention of the worldwide web and several medical technologies — instead of relying on the brilliance of its research to make the case for funds.

Many observers say that since the collapse of the Superconducting Super Collider project in Texas in 1993, high-energy physics has struggled to regain prestige in Washington.

Areas of physics that were once far less fashionable have won more support — solid-state physics, for example, will benefit from President Clinton's recent nanotechnology initiative. But the budget for high-energy physics has been frozen for several years, causing substantial staff reductions at the two largest particle-physics laboratories, Stanford Linear Accelerator Center (SLAC) in California and Fermilab in Illinois.

A meeting of HEPAP at Fermilab last month was told that the two laboratories are facing a \$50 million shortfall in the next financial year, under the Department of Energy's budget plan. Last summer, the laboratories — aided by the fact that Fermilab is



"NEUTRINO FACTORY"
IN RACETRACK SHAPED RING...

in the district of Dennis Hastert (Republican, Illinois), the speaker of the House of Representatives — persuaded Congress to add some \$18 million to the budget proposed for them by the administration.

Leaders of the programme are hoping, meanwhile, that the study requested by Rosen will help to provide direction for the field. "It will be a unifying process for the

community," predicts Jonathan Dorfman, director of SLAC.

The 1998 study said that the top priority was to operate existing facilities at SLAC and Fermilab. It added that research and development work should continue on a proposed electron-positron collider, the Next Linear Collider (NLC), in order to complete a conceptual design for it, and that extended research programmes should investigate two other machines, a muon collider and a very large hadron collider.

Some sources in the high-energy physics community say that Rosen wants the new study to sharpen the focus of this agenda by calling explicitly for the construction of the NLC. But not everyone has rallied behind the NLC. Some physicists believe that, at an estimated cost of about \$5 billion, it is too expensive to be built in the United States in the foreseeable future, and that the community should be exploring other options instead.

"The high-energy physics programme will go the way of the fusion programme if it pursues the NLC as its best bet," warns John Peoples, former director of Fermilab. The

UK science spend 'not sensible'

London

Increases in the British government's spending on fundamental science are being funded by cuts to research in more directly applicable areas, a cross-party group of members of parliament (MPs) warned this week. They said that "a policy of robbing Peter to pay Paul is neither sensible nor sustainable".

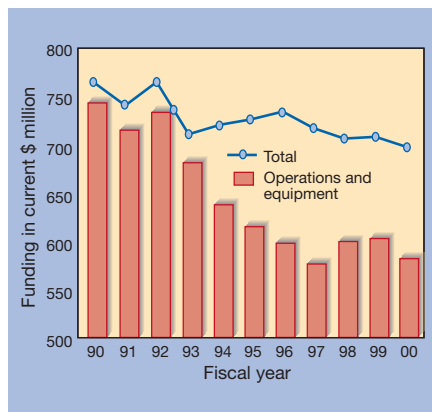
A report from the House of Commons select committee on science and technology points out that civil departments, such as the Ministry of Agriculture, Fisheries and Food (MAFF) and the Department of Health, have suffered a long-term decline in support for their research spending. This is in spite of their role of informing government policy, addressing gaps in the science base and responding to short-term research needs.

The committee says that although public confidence in agriculture is at an all-time low and the farming industry is in crisis, MAFF's research and development budget has been cut by ten per cent in the past three years.

The committee says this decline should be halted or reversed. It also recommends that the minister for science be elevated to Cabinet level. At present science is represented within the Cabinet by the Secretary of State for Trade and Industry, Stephen Byers.

The lobby group Save British Science wants a Ministry of Science — something the cross-party group of MPs does not agree with. But Peter Cotgreave, director of Save British Science, said elevating the position of the science minister would be "a decent compromise".

Natasha Loder



Fading funds: US financial backing for high-energy physics has slowed down significantly.

► fusion research programme in the United States lost one-third of its budget in 1995 after Congress decided that US involvement in the International Thermonuclear Experimental Reactor was too expensive to pursue.

Since the 1998 study, physicists at Fermilab and elsewhere have been developing the

case for a 'neutrino factory', which would generate an intense beam of neutrinos from the decay of muons circulating in a race-track-shaped ring.

Such an experiment, which would cost about a billion dollars, could build on current efforts in Japan, the United States and Europe to detect neutrino oscillation, and therefore measure the mass of the neutrino. Some advocates of the neutrino factory see it as a realistically priced, intermediate step towards a muon collider that could eventually be built at Fermilab.

"There are a lot of people worrying about their own survival, and they are not prepared to take the NLC on as a major thrust," says Lubell. One laboratory official goes further, saying that the current budget crunch "has created a crisis mode of thinking, close to a feeling of panic" in the field which, the official adds, is not conducive to reaching a consensus on future projects.

"It's a very difficult situation," says Nick Samios, former director of the Brookhaven National Laboratory. "Our credibility is not

high, because of the Superconducting Super Collider debacle." Samios supports the NLC construction, but notes that current plans for a collider with an initial energy of 500 GeV, upgradable to 1 TeV, falls well short of a previous assumption by US physicists that such a machine should aim for an energy of 1.5 TeV.

Whatever its eventual specification and site, US physicists are in agreement that the NLC will have to be a truly international project. The \$20 million annual research programme into the NLC, which is centred at SLAC, has been closely coordinated with a similar programme at the High Energy Accelerator Research Organization (KEK) in Japan.

At a meeting in London earlier this month, the global science forum of the Paris-based Organization for Economic Cooperation and Development discussed the formation of a high-energy physics workshop that could provide the necessary framework for the international construction of machines such as the NLC.

Colin Macilwain

'Saturation screen' lets zebrafish show their stripes

London

Over the next year, a mammoth project involving some 17 million tropical fish is expected to yield a treasure trove of data for researchers working on the genetics of vertebrate development.

Artemis Pharmaceuticals, a company based in Cologne, announced last week that it has joined forces with Christiane Nüsslein-Volhard of the Max Planck Institute for Developmental Biology in Tübingen to conduct a zebrafish 'saturation mutagenesis screen'.

The project, dubbed the Tübingen 2000 Screen, is already under way, and will cost around US\$10 million. Its goal is to apply chemical mutagens to zebrafish sperm to introduce point mutations randomly throughout the genome.

Researchers will breed fish using the treated sperm, and screen for those that show developmental abnormalities. From these animals, geneticists can then try to clone the mutated gene. Applying this technique on a sufficiently large scale should make it possible to mutate every important developmental gene — hence the term 'saturation' mutagenesis.

Nüsslein-Volhard shared a Nobel prize in 1995 for her work in developing saturation mutagenesis in the fruitfly *Drosophila*. By this time, she had also begun to apply the technique to the zebrafish (*Danio rerio*) to investigate the function of genes that control the development of vertebrates. This popular aquarium fish is ideal for

developmental studies, as its embryos are transparent, making it easy to spot abnormalities in internal structures.

Nüsslein-Volhard's previous zebrafish screen produced more than 1,000 mutants. A parallel screen — conducted by a team led by Wolfgang Driever, now at the University of Freiburg — produced half as many mutants again. Mutants from both screens were described in a special issue of the journal *Development* (see *Nature* 384, 514; 1996). "A lot of interesting mutants were found, and it has been possible to start to characterize some of those molecularly," says Phil Ingham, a developmental biologist at the University of Sheffield.

But the original screens failed to mutate every important zebrafish developmental gene. "We probably didn't get to saturation," says Nüsslein-Volhard, who is a co-founder of Artemis. In addition, some interesting mutants may have been overlooked because researchers didn't spot subtle developmental abnormalities in every single organ and physiological system.

This time, researchers with different expertise will examine the same fish to ensure that fewer mutants get overlooked. The screen will be twice the scale of Nüsslein-Volhard's previous project. Peter Stadler, chief executive officer of Artemis, says the project will employ about 50 people.

Scientists at Artemis will mainly screen for abnormalities giving information on gene function that could provide leads for drug development. These include



Variety shoal: zebrafish with the *mini fin* gene, one of a host of mutants known in the species.

abnormalities in the production of cartilage and bone, and in the cardiovascular system.

Nüsslein-Volhard's team, together with researchers from elsewhere in Germany, Britain and the United States, will look for mutants yielding fundamental insights into the processes of development. These will include fish with abnormalities in the formation of the eyes and the neural crest — paired strips of cells that form early in development, which give rise to a variety of tissues including the central nervous system.

Artemis will have six weeks to patent any discoveries. After this period, results will be made available to other scientists in the normal way, says Stadler.

Peter Aldhous

► <http://www.artemis-pharmaceuticals.de/html/topic2000.html>

► http://www.eb.tuebingen.mpg.de/abt.3/home_text.html

Letter fuels South Africa's AIDS furore

Cape Town

International controversy over the South African president Thabo Mbeki's views on the nature of AIDS deepened last week, when the text of an outspoken letter sent earlier this month to US President Bill Clinton, UN secretary-general Kofi Annan and the heads of state of Germany, France and the United Kingdom became public.

In the letter, Mbeki gives the strongest evidence yet of his sympathies for those who argue that AIDS may not be caused by HIV.

Ironically, earlier in the week James Wolfensohn, president of the World Bank, had promised that there would be "no limit" to the funds available for combating AIDS in the developing world. In his statement, delivered during the annual World Bank meeting in Washington, Wolfensohn singled out sub-Saharan Africa as being in particular need.

The publication of the letter coincided with the release of South Africa's AIDS statistics for 1999, which confirm that nearly ten per cent of its population — about 4.2 million people — is infected with HIV.

Mbeki's letter was passed to *The Washington Post* by a senior US official, following a meeting in Atlanta the previous week between South Africa's health minister Manto Tshabalala-Msimang, its ambassador to the United States, Makate Sisulu, and Sandra Thurman, director of the White House office of national AIDS policy.

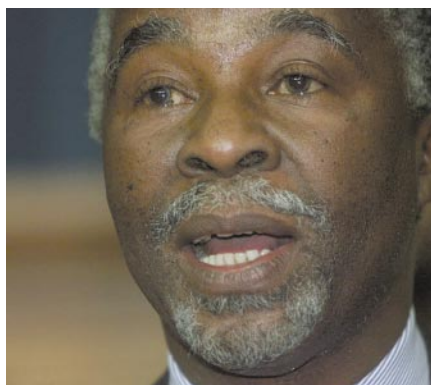
In the letter, Mbeki states that he is "convinced that our urgent task is to respond to the specific threat that faces us as Africans. We will not eschew this obligation in favour of the comfort of the recitation of a catechism that may very well be a correct response to the specific manifestation of AIDS in the West."

Referring to 'dissident' AIDS researchers who believe that HIV does not cause AIDS, he continues: "The scientists we are supposed to put into scientific quarantine include Nobel prizewinners, members of academies of science and emeritus professors of various disciplines of medicine!"

In the South African parliament last Wednesday, Jacob Zuma, deputy president and chairman of the National AIDS Council, denied that Mbeki had at any stage said that he challenges the view that HIV causes AIDS — or the contrary.

"During the last decade and a half and more, a heated debate amongst scientists and others relating to this question, has been going on," said Zuma. Rallying to the president's position, he continued: "We should not, and we will not, leave any stone unturned, even if this means including the views of the so-called 'dissidents'."

Zuma said that an international panel being set up by the South African government to advise on an appropriate national



Mbeki: wants to hear views of AIDS 'dissidents'.

policy on AIDS included "all points of view in the debate and is constituted of the most eminent world scientists who can help to ensure that we understand HIV/AIDS correctly and therefore respond to it correctly".

But the decision to set up the panel, and the way it is being done, has itself come under criticism. Although the panel is due to meet for the first time in South Africa next week, its members have not yet been announced. Two medical researchers, both linked to the dissident movement, are the only people who have confirmed their participation.

One is Gordon Stewart, professor emeritus of public health at the University of Glasgow, who became well known in the

1970s for his opposition to whooping-cough vaccine. The other is Sam Mhlongo, professor of family health at the Medical University of South Africa, north of Pretoria.

In 1997, Stewart co-authored an article with Eleni Papadopoulos-Eleopoulos of the Royal Perth Hospital, Australia, and other 'dissidents' in the journal *Current Medical Research and Opinion*. A later article in the same journal by Papadopoulos-Eleopoulos and others on the pharmacology of AZT, which claimed that the drug was unacceptably toxic, appeared to be a major influence in Mbeki's refusal to sanction state provision of this drug to prevent mother-to-child transmission of HIV (see *Nature* 402, 3; 1999).

During the debate in parliament last week, Mike Ellis, a member of the Democratic Party, said that Mbeki seemed so defensive about AIDS that he was placing sound judgement and rational thinking at risk. Kobus Gous, health spokesman for the New National Party, accused the president, through his contact with the dissidents, of having "given a podium to discredited scientists, false hope to AIDS sufferers and created doubt in the public mind".

But support for Mbeki's stance came from an unusual quarter when the right-wing Boerestaart Party applauded the president's efforts to "investigate the biggest hoax in the twentieth century". **Michael Cherry**

Unesco backs 'science for debt' plan

London

Koichiro Matsuura, the new director-general of the United Nations Educational, Scientific and Cultural Organization (Unesco), promised this week that his organization will "spare no effort" in backing poor countries prepared to draw up proposals for swapping debt relief for support for science.

Matsuura was speaking at a meeting of African leaders held on Monday (24 April) in Abuja, the capital of Nigeria, to launch an anti-malaria campaign. "Without a scientific capacity of its own, Africa will not be able to tackle and overcome its endemic diseases," he said.

The idea that African countries in particular should develop a strategy under which money earmarked for debt repayments could be channelled into supporting science was endorsed last summer at the World Conference on Science in Budapest, organized jointly by Unesco and the International Council for Science (see *Nature* 400, 8; 2000).

After the Budapest meeting, Unesco commissioned a report from two former



Matsuura: 'will spare no effort'.

officials of the World Bank on how this could be achieved. The report describes how scientists can integrate their demands into the policy packages put together by their governments under the Heavily Indebted Poor

Countries Initiative, a plan drawn up jointly by the World Bank and the International Monetary Fund.

The report says, for example, that it is "essential" to develop an advocacy strategy within a country for supporting science and technology, given that "the bargaining for distribution of resources is an in-country process".

In his speech, Matsuura said the report shows that it is feasible to integrate a science-funding component into negotiations on debt relief, "whatever stage [they] have reached". He added that "debt relief for science should be seized on as an opportunity to strengthen local capacities in basic and applied research". **David Dickson**

Genetic diversity project fights for its life...

Vancouver

A combination of political and public controversy has reduced the size and scope of the Human Genome Diversity Project (HGDP), a bold scheme to collect, store and analyse DNA representing the world's ethnic diversity.

This ambitious goal has now become a disparate set of collection projects, funded from a variety of sources. A DNA repository and database — two key components — seem unlikely to happen without more funding and broader political support.

Even some of the project's strongest supporters are somewhat pessimistic about its prospects. Hiraku Takebe, professor at the Atomic Energy Research Institute at Kinki University in Osaka, Japan, told the Human Genome Organisation (HUGO) meeting in Vancouver earlier this month that he fears the project is "almost aborted". HUGO began exploring the HGDP formally in 1992, but in recent years hasn't aggressively pursued it, says Takebe.

He adds that, without the inclusion of genomic data from populous countries such as India and China, the Human Genome Project — which is constructing a composite genome by sequencing DNA from several US individuals from a variety of ethnic backgrounds — could more aptly be called the 'Western Nation Genome Project'.

But even without HUGO's backing, the HGDP faces serious obstacles. Some countries, such as New Guinea, are now demanding a fee for each DNA sample that researchers take. Others, such as India, forbid the export of DNA completely.

What's more, says Takebe, other projects with related goals, such as the SNP (single-nucleotide polymorphism) Consortium, could upstage the HGDP's efforts. The consortium, which aims to define the single base-pair gene differences often associated with disease susceptibility or response to drugs, has received US\$50 million from private and public sources.

The HGDP is asking broader questions about other kinds of genetic diversity. But unlike the commercially driven SNP Consortium, the HGDP has always struggled to attract support.

After population geneticist Luigi Luca Cavalli-Sforza of Stanford University conceived the idea, HUGO formed a committee to pursue it, with Cavalli-Sforza as its chair — a position he still holds. But the project has met with ambivalence at best from governments and funding agencies. HUGO lacks the resources to support large projects itself.

The United Nations Educational, Scientific and Cultural Organization failed to endorse it in 1995, partly because of the controversy surrounding the project's collecting



At the margin: ethnic diversity has been omitted from the Human Genome Project.

of genetic samples from populations around the world. And in 1997, a report from the US National Academy of Sciences recommended the US National Science Foundation and National Institutes of Health not to fund the project (see *Nature* 377, 373; 1995 & 389, 774; 1997).

"Our lack of funding is a huge issue," says Henry Greely, professor at Stanford Law School and a spokesman for the project. But he denies that it is killing the programme. "Rumours that the HGDP has died, or is in a comatose state, are exaggerated," he says.

One of the most frequent criticisms of human genetics research is that it takes

genetic information from populations without giving enough back. Earlier this month, HUGO moved to address this concern with a statement on 'benefits sharing'. But HUGO's new policy is purely advisory, and differs substantially from guidelines drafted by the HGDP (see below).

The current HUGO leadership appears to be ambivalent about the HGDP — at least as an ambitious, centralized effort. HUGO president Lap-Chee Tsui, geneticist-in-chief at the Hospital for Sick Children in Toronto, says the programme doesn't have clearly defined goals and questions.

Tsui says that the HGDP has been framed variously as a search for disease genes, a way to track ethnic migration, and a means to identify disease susceptibility and prevalence in different populations. Greely agrees that the project potentially addresses many research questions. But he disagrees that this is a weakness.

Not everyone is pessimistic about the project's prospects. Walter Bodmer, a geneticist at the University of Oxford and former HUGO president, predicts that, as more scientists become aware of the uses of data on genetic variation, support for the HGDP will grow. "Population variation is only going to become more valuable," says Bodmer. "I think the tide is turning."

Cavalli-Sforza says that the work he and Bodmer have done shows that diversity research is healthy. Scientists in other countries, including India and China, are collecting the same sort of data. "Whether they call themselves HGDP or not really doesn't matter to me," Cavalli-Sforza says. **Paul Smaglik**

... as companies are urged to share benefits

Vancouver

Sharing doesn't always come easily. But it may be necessary to generate research that relies on obtaining genetic samples from people from a variety of geographic locations and ethnic backgrounds.

The Human Genome Organisation (HUGO), at its annual meeting in Vancouver earlier this month, released guidelines encouraging pharmaceutical companies and others to donate 1–3 per cent of their annual net profit to health infrastructure and humanitarian aid. If widely adopted, these guidelines could encourage more countries to

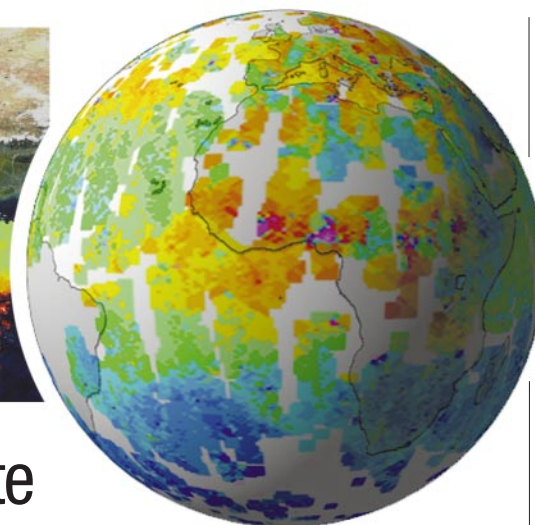
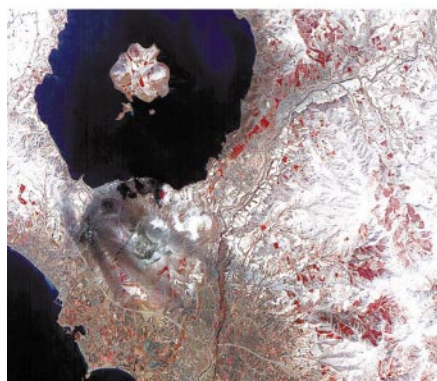
participate in efforts such as the Human Genome Diversity Project (see above).

But adoption is unlikely to come easily, admits Bartha Knoppers, chairperson of HUGO's ethics committee, which wrote the guidelines. "At the beginning, it may sound utopian or naive, but if you don't bring it up, the issues don't advance," says Knoppers, a law professor at the University of Montreal.

Many subtleties complicate the matter. For instance, HUGO's guidelines emphasize helping an entire country or community, whereas HGDP's guidelines specifically favour rewarding those who cooperate.

In some cases, this is straightforward. For example, people with Tangiers disease — a rare genetic condition resulting in lower levels of HDLC, or 'good' cholesterol — could be easily rewarded for their involvement if a drug is developed, says Kare Berg, a geneticist at the University of Oslo.

But tracking down everyone who has ever participated in a cancer trial that may have contributed to a successful therapy years after their participation is more complicated, says Dorothy Wertz, a sociologist at the Eunice Kennedy Shriver Center in Waltham, Massachusetts. **P.S.**



NASA's new environmental satellite shows its sensitive side

Washington

Beaming scientists last week showed off the first images from NASA's Terra satellite. The flagship of the \$11 billion Earth Observing System, Terra is dedicated to long-term monitoring of the home planet's surface, ocean, ice and atmosphere.

All five of Terra's instruments have checked out "in great shape" since the satellite's December launch, says NASA's head of Earth science, Ghassem Asrar. Although it is likely to take months for Terra to produce

significant scientific results, the instruments are already showing their great sensitivity.

For example, the Japanese ASTER (advanced spaceborne thermal emission and reflection) instrument caught the aftermath of Mount Usu's volcanic eruption on 31 March (see top left). The three dark streaks in this visible-infrared image are ash deposits.

The US-built moderate-resolution imaging spectroradiometer (MODIS), aside from tracking vegetation changes on the land, is unique among satellite sensors in its ability

to monitor the health of plankton in the ocean by measuring fluorescence (see top centre). This MODIS view over the Arabian Sea shows the highest amounts of photosynthetic activity in blue and lower ones in red.

The Canadian MOPITT (measurements of pollution in the troposphere) sensor maps atmospheric carbon monoxide at levels of one part in ten million (see top right). The red and violet areas over western Africa in this image show high carbon monoxide levels caused by seasonal fires.

Tony Reichhardt

Japan gears up for growth in genomics

Tokyo

Plans for a high-speed genome sequencing centre, announced earlier this month by biotechnology company Takara Shuzo, are widely being seen as evidence of an increased willingness by Japanese industry to invest in genomics and post-genomics research.

The new centre, which represents an investment of US\$60 million, will be located in Kusu-cho, Mie Prefecture. Called Dragon Genomics, the centre will have a sequencing capacity of about half the peak capacity of the US company Celera Genomics, making it the largest sequencing facility in Asia.

"This is the first corporate genome sequencing centre in Asia with the capacity of doing genome-size sequencing efficiently," says Makoto Moriguchi, director for sales and marketing in Takara's Biomedical Group, and business development manager of the project.

Takara Shuzo has chosen sequencing equipment developed jointly by scientists at the Research Institute for Physical and Chemical Research (RIKEN) in Tsukuba, near Tokyo, and the instrument manufacturer Shimadzu. To cut costs, sample prepa-

ration will be undertaken by Takara Biotechnology, a division of Takara Shuzo in Dalian, North China. Functional analysis of the genes will be done at the company's research centre in Shiga, Japan.

According to Moriguchi, Dragon Genomics will initially offer contract sequencing services to pharmaceutical and food-processing companies both in Japan and internationally. But Takara also plans to sell sequencing data of important microorganisms, and it will seek patent protection for genes encoding useful proteins. About a fifth of the new centre's capacity will be dedicated to marine biotechnology.

"We do not intend to target the human genome, since this is not an area where we feel we can compete any more. But there are many other areas where much remains to be done," says Moriguchi.

Following a surge in public funding for genomics and post-genomics within this year's government research budget, equipment manufacturers Hitachi and Shimadzu have also announced their intention to expand contract sequencing facilities.

According to Tetsuo Ichikawa, managing officer of Shimadzu's analytical instruments

division, the company plans to provide sequencing services to national research centres and universities.

But while Shimadzu sees contract sequencing services mainly as part of a strategy to sell its equipment, the electronics giant Hitachi has taken the unusual step of setting up an independent life-science group. Hiroya Taguchi, president and chief executive officer of Hitachi's Life Science Group, says the main focus of the new group, which was set up formally last September, will be on sequencing and database services for pharmaceutical companies and public organizations.

Taguchi says that Hitachi is "several years behind US and European companies" and will initially target only Japanese clients. But, he argues that the company's experience in information technology and high-performance computing means that it is well positioned to become a global player in genome informatics.

According to Moriguchi, Takara Shuzo also plans to install a high-performance computing facility at its Shiga research centre. But Moriguchi concedes that the company lacks experience in bioinformatics and is currently seeking a partner.

Robert Triendl

New Zealand GM inquiry will cast a wide net

Sydney

New Zealand's new Labour-led coalition government has set up a royal commission into the implications of modifying genes. The commission's main purpose is to reconcile a bitter debate within the country over GM crops, but it is also expected to have international implications.

The government has imposed a 'voluntary moratorium' on applications to release and field-test genetically modified organisms (GMOs) until the commission reports, and a total ban on releases "involving reproductive material".

The status of the inquiry is believed to give the four-member panel an independence and power to call and cross-examine witnesses not enjoyed by inquiries into transgenic organisms in other countries.

Sir Thomas Eichelbaum, who retired as the country's Chief Justice last May, will chair the commission. He expects the proceedings to be a mix of testimony given under oath — with legal representation — and less formal discussion. "Our conclusions will never be the definitive word," he says.

Although its remit is to report on strategic options and to recommend changes to legislative, regulatory, policy or institutional

arrangements in New Zealand — whose economy is dependent on agriculture — the inquiry will be international in scope.

Opponents of genetic engineering are preparing to press their case at the inquiry. Doug Parr, Greenpeace's chief scientific adviser in Britain, says that his organization is willing to participate. "No country has the monopoly of expertise in this issue," he says. Multinational biotechnology companies are also expected to take part.

"All sides are welcome and we are providing resources from a budget of NZ\$4.8 million (US\$2.5 million) to enable submissions from overseas, in person or by video link," says Marian Hobbs, New Zealand's environment minister. She believes that the style of the inquiry will "help to dispel distrust by sharing one vocabulary".

Hobbs announced the inquiry in a week when two breaches of the regulations on GMOs were revealed. One was by a laboratory at the University of Otago's Medical School in Christchurch, where researchers allegedly extracted tissue from the lizard-like tuatara and were constructing a DNA library, without full permission from the regulatory authority or local Maori people, for whom the animal has spiritual significance.



Eichelbaum: accepts that his commission's conclusions "will never be the definitive word".

The chair of the university's biological safety committee, George Petersen (who is also acting president of the Royal Society of New Zealand) himself raised the alarm, declaring "sadness for the whole scientific community". But Petersen believes that "by publicly confessing before the regulators or the media uncovered it, we earned good points for owning up".

The commission's other members are Jean Fleming, a researcher at the University

of Otago in molecular reproduction and endocrinology; Richard Randerson, an Anglican priest with a strong record in social issues; and Jacqueline Allan, a medical practitioner of Maori descent.

The opposition National Party has attacked the lack of any panel members with industrial experience. But Hobbs rejects this, saying that “the production community will be welcome to contribute”.

University scientists appear reconciled to the ban on GMO release, as they have escaped the restrictions on laboratory work sought by the Green Party. Researchers can still argue for exemptions for non-commercial work, such as the control of possums. But Greenpeace and other activists have described the moratorium as contradictory and voiced concerns over “the risk of irreversibility”.

The Crown Research Institutes, set up by the previous government and funded through contract work from both the public and private sector, have angered conservationists by pressing for expanded field trials. But science minister Pete Hodgson has told the institutes to toe the government’s line.

Spokesman and head of HortResearch, Ian Warrington, says that the restrictions “will cause the termination of some research”. Michael Dunbier, chief of the Crop and Food Research Institute, says the ban could cause scientists to leave New Zealand. **Peter Pockley**

German research agency ‘doesn’t stifle creativity’, say 1,600 scientists

Munich

More than 1,600 German scientists have rallied to the defence of the country’s main grant-giving agency for basic research, the Deutsche Forschungsgemeinschaft (DFG). Their action is a response to allegations that the agency is reluctant to fund research outside the scientific mainstream.

In today’s *Nature* (see page 922), the scientists defend the DFG against criticism from parts of Germany’s scientific community recently reported in German newspapers (see *Nature* 404, 217; 2000).

The complaints, they argue, are unrepresentative and largely unjustified, given the DFG’s generally “unbiased support for creative, high-quality research and its programmes for young scientists”.

The letter was drafted by Reinhard Jahn and Herbert Jäckle, directors at the Max Planck Institute for Biophysical Chemistry in Göttingen. It was mailed to a number of randomly chosen scientists, with a request to sign it and forward it to colleagues.

The initiative was well received — albeit mainly by senior scientists. Only about a

quarter of the 1,600-odd signatories are young scientists such as PhD students or postdoctoral researchers. Most are institute heads or other well-established researchers.

But Jahn, 48, an experienced DFG referee and a winner of the agency’s DM3 million (US\$1.4 million) Leibniz prize for his research on biological membranes, says that “the time has come for the whole community to positively affirm the DFG and to protect it against damage”.

Jahn says he decided to initiate the chain letter after reading last month’s *Nature* article. “Such negative reports about the DFG are grist to the mill of those who would like to increase political influence on our self-governing agency,” he says.

Jahn returned to Germany in 1997 after six years at the Howard Hughes Medical Institute in Yale. He challenges the claim that the DFG is less efficient than US or UK funding agencies. “Compared, for example, with the procedures of the European Commission, the DFG is certainly capable of funding the best research in a highly efficient manner,” he says. **Quirin Schiermeier**

Drosophila genome contaminated with human sequence

Washington Computational biologists at the National Center for Biotechnology Information (NCBI) have discovered about 150,000 bases of human DNA mixed in with the genomic sequence of the fruitfly *Drosophila* compiled by Celera Genomics of Rockville, Maryland, and the Berkeley *Drosophila* Genome Project (BDGP).

Scientists working with the data were downplaying the significance of the contamination last week. David Lipman, NCBI director, says that the figure should be seen in the context of the approximately 180 million bases of the fly genome.

Lipman adds that the errant data were in a part of the database holding 'unassigned scaffolds', not the labelled part of known fly chromosomes. "We found no evidence of any contamination in the main body of fly sequence data," he says.

Researchers at Celera and the BDGP had previously pointed out that some of the unassigned sequence might contain foreign DNA. They have requested a correction in *Science*, which published the *Drosophila* sequence last month. Gerald Rubin, leader of the BDGP, has been quoted as describing the contamination as "trivial".

Russia and North Korea sign cooperation pact

Moscow Russia and North Korea signed an agreement on science, education and culture last week, intended to increase collaboration between researchers in the two countries. They have agreed to promote exchanges between their research institutions, especially the national academies of sciences, and to expand cooperation between their higher educational institutions.

There will also be regular exchanges of students and teaching staff, aimed at improving their language skills. The agreement was signed by Vladimir Dorokhin of the Russian Foreign Ministry, and Choe Joung Hun, the head of the department of cultural exchanges at the North Korean committee for foreign ties.

Indian scientists seek help in tackling fraud

Hyderabad, India Indian scientists have decided to ask the government's Central Vigilance Commission (CVC) to establish an office to deal with allegations of scientific fraud and misconduct. The decision was taken last week at a symposium in Hyderabad on ethics in science, organized by the Society for Scientific Values (SSV).

A government body, the CVC has gained

admiration for its anti-corruption policy of naming suspect government officials and politicians on the Internet. At the Hyderabad symposium, it was urged to treat fraudulent researchers similarly. The nine-year-old SSV has unearthed several such cases, but it lacks the clout to punish offenders.

Neuroscientist carries off inaugural Hungarian prize

Budapest The Bolyai Prize — recognizing Hungarians achieving international success in scientific research, development, teaching or applications — has been awarded for the first time. Tamás Freund, head of the Department of Neurosciences at Pázmány Péter Catholic University in Budapest, received the prize — worth US\$50,000 — for his research on the brain.

The Bolyai Prize Foundation was established by four Hungarian entrepreneurs to support the social acceptance of science, and to encourage similar charitable ventures by Hungary's 'new capitalists'. The prize will be awarded every two years.

UK court undecided on GM activists

London The first court case of criminal damage to a research trial of genetically modified (GM) crops ended last week when a British jury failed to reach a verdict. As a result, 28 protesters from the environmental group Greenpeace, led by the group's executive director and ex-government minister, Lord Melchett, were discharged from court.

The protesters had pleaded not guilty, arguing that they were acting to protect other crops and the environment from genetic contamination by GM maize pollen. A spokesperson for the Department of Environment, Transport and the Regions said there were "no implications yet" for this year's enlarged trials, but described the lack of a verdict as "not helpful".

Decision on NIH head put back until after election

Washington Ruth Kirschstein, acting director of the US National Institutes of Health (NIH), is to stay at the head of the agency until after this year's US presidential election, allowing the next president to nominate a new director.

The Department of Health and Human Services, which oversees NIH, had considered nominating Gerald Fischbach, director of the National Institute of Neurological Disorders and Stroke, for the post. But department officials and Fischbach recently decided against pursuing the nomination, because, even if he were to have received rapid Senate approval, he would have held the position for only a few months before the arrival of a new administration.

Lyon will house WHO disease watchdog

Paris The World Health Organization (WHO) has announced plans to establish its world centre for epidemiological surveillance and alert in Lyon, France. The centre will coordinate with 200 regional offices, and will focus on collaborating with countries in the developing world on the diagnosis of transmissible diseases, notably HIV, Creutzfeldt-Jakob disease and the Ebola virus. The WHO facility will be the first of its type in Europe: there are similar centres in Atlanta, Georgia; Fort Detrick, Maryland; and South Africa.

Stanford fund for graduate students up and running

San Diego Stanford University has achieved its target of raising \$200 million for a fellowships programme to support promising graduate students. According to university officials, the programme will be able to provide \$31,500 a year for three years — \$13,000 for tuition plus a \$18,500 stipend — for at least 300 students. Alumni, foundations and corporations contributed towards the endowment, which will enable graduate students to conduct research in all fields.

DNA confirms dead child as French prince

Paris DNA tests have confirmed that a child who died on 8 June 1795, in the Prison du Temple in Paris, was Louis XVII, the son of Louis XVI and Marie Antoinette (pictured with her children below). Researchers' conclusions, announced last week, are based on tests done in Belgium and Germany on the boy's heart tissue and hair from Marie Antoinette, as well as tests on two of her sisters and two of her descendants.

But the results are disputed by the Institute of Louis XVII, which believes that the imprisoned dauphin was replaced with another child. The Institute has asked laboratories in Nantes, France, and the United States to do more tests.



BRIDGEMAN ART LIBRARY

Taiwan's kingmaker chemist

Nobel chemistry laureates don't normally enter the hurly-burly of politics — but Lee Yuan-tseh did. David Cyranoski meets a reluctant political hero.

AP

As the new head of Taiwan's leading scientific academy, Lee Yuan-tseh should feel comfortable in the shoes he has to fill — they're his own.

Last month, Lee resigned as president of Academia Sinica after throwing his weight behind Chen Shui-bian, the reformist presidential candidate who won the 18 March election. The next ten days were a whirlwind, as Lee was offered the chance of leading Chen's government. Now, having declined the post, Lee is back at his old desk.

Lee is Taiwan's best-known scientist, having shared the 1986 Nobel Prize in Chemistry for his studies on elementary chemical processes. He is also a popular public figure — mentioning his name can draw an enthusiastic thumbs-up from Taipei's taxi-drivers. In a close-run poll, commentators say that Lee's endorsement was instrumental in tipping the scales in favour of Chen's Democratic Progressive Party (DPP).

It was a controversial move. The head of Academia Sinica reports to Taiwan's president. So in backing Chen, Lee was opposing his boss, outgoing president Lee Teng-hui — hence his resignation. But, says Lee, he never wanted to be premier: "I have no interest in politics." The only reason Lee delayed his announcement, he adds, was to give Chen time to find an alternative candidate.

Despite this professed lack of interest in politics, Lee has helped to reshape Taiwan's political landscape. Chen's victory ends 55 years of government by the Kuomintang party, and brings in a party that has called for Taiwan to officially become an independent republic. This is anathema to Beijing, which remains committed to a united China, with the island province under its control.

Lee has arranged academic conferences to bring mainland scholars to Taiwan, and he hopes to continue to strengthen scientific links with the mainland. In this, he faces a challenge, as his alliance with Chen has tarnished his once-pristine image in Beijing. However, in the run-up to the election, the DPP stepped back from its pro-independence rhetoric. And in a move calculated to appease the People's Republic, Chen has nominated Tang Fei, defence minister in the outgoing government, as premier.

Lee will have a key role in building relations with Beijing. At Chen's request, he will head a multi-party group to reach a consensus on Taiwan's strategy in talks with the mainland. "Most do not object to the 'one



Allies: Lee, and president-elect Chen (above left).

China' concept," Lee says. "The problem is deciding what kind of China that will be."

Lee returned to Taiwan from the University of California at Berkeley in 1994. During his six-year tenure at the Academia Sinica, which runs a network of 24 research institutes, government research spending surged. Annual increases have averaged close to ten per cent, well above the overall growth in public spending. Scientists say Lee's influence has been crucial to this policy, which has helped to plug a scientific brain drain.

Although they are delighted to have Lee back at the helm of Academia Sinica, Taiwan's scientists now wonder whether his political sojourn will have any repercussions. The Kuomintang retains a majority in the legislative Yuan, which votes on the academy's budget. But the party is splintering, and does not seem likely to pick a fight with Lee. Sunney Chan, vice-president of Academia Sinica, notes Lee's "unique ability to appeal to the business sense of both those in the government and in industry".

Lee has also appealed to scientists' entrepreneurial sense. In his previous tenure at Academia Sinica, Lee launched a Technology Transfer Office that urges researchers to take advantage of their intellectual property rights. This broke even after three years, instead of the projected five.

In a close-run poll, Lee's endorsement was instrumental in tipping the scales.



He now wants to launch a programme to help start-up businesses in Taiwan's nascent biotech sector, and in areas such as advanced materials. But he remains a supporter of basic science, such as the joint project, with the Smithsonian Astrophysical Observatory in Cambridge, Massachusetts, to develop an array of radio telescopes in Hawaii.

Lee also hopes to use his influence to promote a more civil society. The reason he sided with Chen, he says, was that the DPP promised to root out corruption — known to the Taiwanese as 'black gold'. Lee is tackling this issue as head of the non-profit Society for Community Empowerment, which promotes activities from recycling to music festivals. It also gave aid to areas hit by last year's earthquake.

Pressed on why he declined the premier's job, Lee tells of how he was playfully chided by the mayor of Jerusalem, for musing on the city's divisions. "That's what's wrong with you scientists," said the mayor. "You want to solve every problem in a short time. Many problems you just have to learn to live with."

Lee, who retains the air of a thoughtful academic, seems content to be back dealing with familiar problems.

David Cyranoski is *Nature's* Asian-Pacific Correspondent

► <http://www.sinica.edu.tw/as/asbrief.html>

ACADEMIA SINICA

Meet the spin doctors . . .

A small band of researchers is plotting a revolution in electronics — one that exploits the spins of electrons, rather than their charges. Philip Ball profiles the emerging field of spintronics.

The last time electronics was reinvented was in 1948, when Ralph Bown, research director of Bell Laboratories in New Jersey, announced “the electrical equivalent of a vacuum tube amplifier.” “We have called it the transistor,” he told the press.

At the time, not even the transistor's inventors, Bell Labs researchers John Bardeen, Walter Brattain and William Shockley realized just how revolutionary their semiconductor device would be. It changed the conceptual landscape of electronics. No longer would electronic engineers think in terms of electrons travelling along metal wires or through a vacuum. Instead, they were able to construct gates and interfaces with which to manipulate the ‘charge carriers’ in semiconductors — electrons and positively charged ‘holes’.

The latest reinvention of electronics may herald a similar conceptual revolution. Some call the field magnetoelectronics. But it is the catchier tag of ‘spintronics’ that best illustrates the new discipline's departure from conventional electronics. Rather than using electrical fields to manipulate a flow of electrons using their charge as a handle, spintronics marshalls electrons through their spin.

Metaphorical marvel

The spin of an electron is a quantum-mechanical metaphor that arises from Paul Dirac's relativistic wave equation, derived in the late 1920s. Spin is a characteristic as fundamental as charge, but has no analogue in the macroscopic world. An electron does not spin like a top, but it can nevertheless be described by a quantum number character-

istic of its intrinsic angular momentum. And because the spin of an electron imparts a magnetic moment, electrons can be manipulated by magnetic fields.

It is still too early to say whether spintronics will sound the death knell for conventional microelectronics, just as silicon devices did for the vacuum tube. But in principle, it could move electronics from a technology grounded in classical physics to one that fully exploits the quantum nature of electrons — perhaps preparing the way for ultrapowerful quantum computers.

Yet even if this young field produces nothing more spectacular than a handful of useful devices, it should have a powerful impact on information technology. Companies including IBM and Motorola are investing heavily. The US Defense Advanced Research Projects Agency (DARPA), the high-tech research arm of the Pentagon, plans to spend “well over US\$100 million” on the field over the next few years, according to Stuart Wolf, who oversees the agency's spintronics projects.

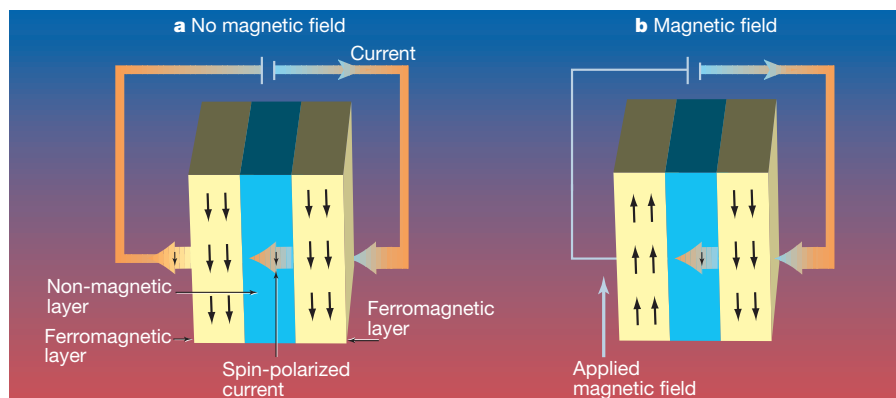
Indeed, magnetoelectronic devices are already in commercial use, as read-out heads for computer hard drives. The market for these devices, which were introduced by IBM in 1997, is now worth around \$1 billion per year. The read-out heads exploit an effect called giant magnetoresistance (GMR), which occurs in multilayer ‘heterostructures’ composed of alternating thin films of a ferromagnetic metal, for example cobalt or iron, and a non-magnetic metal such as copper. The electrical resistance of these structures can be altered profoundly by a magnetic field.



Well read: spintronics has dramatically increased data storage densities in hard drives.

Like all spintronics devices, GMR multilayer structures rely on spin-polarized transport: a flow of current in which electrons are in a spin-aligned state, either spin ‘up’ or ‘down’. This arises naturally in ferromagnets. When such a material is magnetized, the spins of its constituent atoms all align in one direction. Electrons passing through a thin film of the magnetized ferromagnet acquire this same spin bias, creating a partially spin-polarized current.

In GMR hard-drive read-out heads, a current of this type is created when electrons pass through the first magnetic layer. Whether this current passes through the next ferromagnetic layer depends on whether the spins of its electrons are aligned with those of the atoms of the layer. Only if this is so can the current pass freely. In GMR multilayers, the magnetization — and therefore the orientation of the atomic spins — in successive ferromagnetic layers alternates. So the passage of a spin-polarized current is hindered. (Because the spin polarization conferred on a current is only partial in ferromagnetic metals, a small amount current still flows.) But if an external magnetic field is applied, which aligns the spins of all the ferromagnetic layers, the barrier to spin-polarized transport is removed and the resistance drops.



A spin valve in action. a. With no magnetic field, the spin-polarized current can flow. b. When a magnetic field is applied, the spin-polarized current cannot pass through both ferromagnetic layers.

The result is an electronic device that is very sensitive to external magnetic fields. Hard-drive read-out heads based on this principle can register smaller magnetized domains than the previous generation of devices, which used a much more modest magnetoresistive effect in permalloy, an alloy of iron, nickel and molybdenum. Already, this has allowed storage densities of commercial hard drives to be increased by a factor of three. Given recent laboratory findings, Stuart Parkin of IBM's Almaden Research Center in San Jose, whose group pioneered work on GMR read-out heads, has no doubt that it will be possible to improve storage densities a further several-fold.

Memorable events

In the early 1990s, Parkin's group proposed that, for fully spin-polarized currents, a mere three-layer device of this sort can act as a 'spin valve'. These devices comprise a film of non-magnetic metal sandwiched between two films of ferromagnetic material. A spin-polarized current will pass when the magnetization is aligned in the two ferromagnetic layers, but will be totally blocked if the magnetization of just one of them is flipped, for example by a nearby electromagnetic induction loop (see diagram). This is, in essence, the GMR effect taken to its extreme.

Once flipped, the magnetic film in a spin valve retains its direction of magnetization until it is switched once again. This means that the device can act as a memory element. Switching is fast and easy but, unlike the switching in transistors, the device retains its switched configuration when the power is turned off. This means that spintronics could furnish computers with 'non-volatile' magnetic random-access memories (MRAMs), which are not wiped clean by a power failure. They would be particularly valuable on board satellites, where a temporary loss of power can be catastrophic.

Prototype MRAMs consisting of arrays of electrically connected spin valves have been developed by Honeywell, IBM and Motorola. In terms of storage capacity, Honeywell has led the way: for military use, it is producing MRAMs in the megabit range. These are too expensive for consumer applications, but IBM has achieved affordable MRAMs of about 1 kilobit capacity with access times of less than 3 nanoseconds — comparable to those of conventional semiconductor RAMs. These devices all use a subtly different and stronger magnetoresistive effect that depends on 'spin-dependent electron tunnelling'. This is a quantum-mechanical property of electrons that allows them to travel across an interface between two materials even when, according to classical physics, they do not have sufficient energy to do so.

In theory, similar spintronic devices might be used not just for information storage, but also as logic gates — the building



Ohno: pioneering semiconductor spintronics.

blocks from which the circuitry of a microchip is constructed. Not only are spintronic logic gates potentially very fast, but they could also be reconfigured by applying magnetic fields. A prototype device of this type was reported in February by Russell Cowburn and Mark Welland of the University of Cambridge¹. "But no one has demonstrated that you can reliably write and reconfigure a processing function in such a circuit," cautions Parkin. He believes this may prove very difficult to achieve.

Semiconductor sandwiches

Another obstacle for spintronics is that electronics companies are geared up to working with semiconductors, not metals. So one important goal is to make devices using semiconductors that are compatible with existing chip technology. The problem is that the conventional semiconductors used in integrated circuits — primarily gallium arsenide (GaAs) and silicon — are not magnetic. At least, they never used to be. Now, however, several research groups are exploring ways of turning these semiconductors into ferromagnetic materials.

Hideo Ohno of Tohoku University in Japan is one of the pioneers. In 1996, his team showed that ferromagnetic atoms such as manganese can be introduced into GaAs chips fabricated by molecular-beam epitaxy (MBE) — a technique that deposits materials one atomic layer at a time. The solubility of manganese in GaAs is normally very low, but MBE can create a material in which up to 7 per cent of the gallium atoms are substituted for manganese². This alloy behaves as a ferromagnet at temperatures up to 110 K.

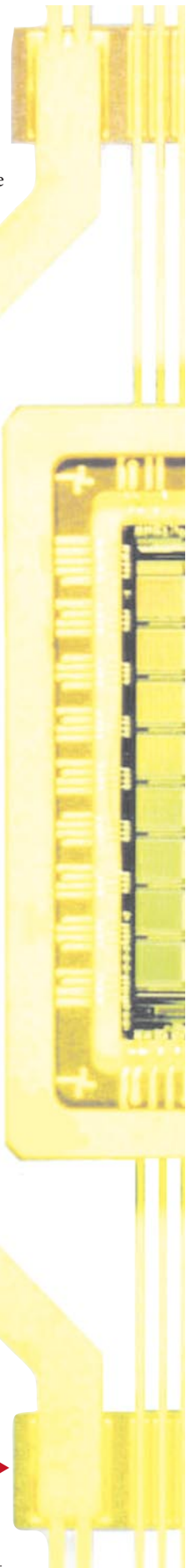
Spintronics will offer completely new kinds of functionality in microelectronics.

Using this doped material, Ohno and his colleagues have made a three-layer device based entirely on GaAs, in which the non-magnetic layer consists of aluminium gallium arsenide. Earlier this month, they reported observing a magnetoresistive effect³. But "it's not giant, yet", says Ohno — the device does not behave as a fully switchable spin valve.

The biggest problem in building such devices is sustaining spin-polarized transport across the interfaces between different materials. Interfaces between semiconductors and ferromagnetic metals, where a double layer of charge carriers forms, are especially problematic. This 'Schottky barrier' lets current flow in only one direction and can scatter charge carriers moving across the interface, which leads to loss of spin polarization. Last year, however, a team led by Mark Johnson of the US Naval Research Laboratory in Washington DC claimed to have injected a spin-polarized current from a nickel-iron alloy into a semiconductor⁴. But the reported efficiency was low, which would make spin valves based on such composite structures extremely leaky.

Injecting spin across the interface between two semiconductors — one magnetic and one non-magnetic — should be easier, because there is no Schottky barrier. And late last year, two groups succeeded in transporting a spin-polarized current across such a junction. Laurens Molenkamp and colleagues at the University of Würzburg in Germany used

Magnetic memory: Honeywell uses spintronics in its RAM chip.

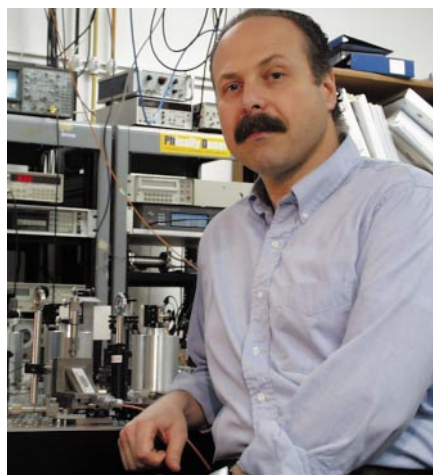


▶ the magnetic semiconductor zinc selenide doped with beryllium and manganese as the spin-aligning ferromagnetic layer⁵; Ohno, working with David Awschalom's team at the University of California, Santa Barbara, used manganese-doped GaAs⁶. In both cases, the researchers passed their spin-polarized current into a GaAs-based light-emitting diode with an efficiency of about 90 per cent. Their success was confirmed by the fact that the emitted light was circularly polarized — a direct consequence of the spin polarization of the current.

These results imply that fully switchable all-semiconductor spin valves should be feasible, and Awschalom cautiously predicts that such devices will be demonstrated within the next few years. He also believes that a semiconductor spin transistor, which would amplify a current rather than merely switching it on and off, will be developed on a similar timescale.

Beyond MRAMs and spin transistors, there is an even more ambitious vision for spintronics: fully exploiting the quantum-mechanical nature of spin. Interest in this possibility has been stimulated by the idea of quantum computing. This offers the prospect of massively powerful information processing by manipulating quantum objects in a 'superposition' of states. Using the language of classical physics, a spin is either up or down, but quantum mechanics means that it can also be held in a superposition that is simultaneously up and down. This makes spin an ideal candidate for the quantum bits (qubits) of quantum computing. Whereas the binary-encoded bits of classical computing have two available states, 1 and 0, qubits have more — greatly expanding computing power.

But processing information with many qubits requires that they are all maintained in a 'coherent' superposition — an interweaving in which none of the qubits is allowed to collapse into its up or down state. This is the



Awschalom: a "genuine quantum electronics" would exploit electrons' wave-like properties.

hard part, because any interactions of the qubits with their environment tends to induce this 'decoherence', losing information.

However, recent studies have raised hopes that it might be possible to sustain spin coherence among charge carriers in solid-state circuits. This is not the same as simply imposing spin polarization on a current, just as polarized light is not the same as a laser beam. Coherent spin carriers are not simply aligned but are coupled in a collective quantum state, described by a single wavefunction.

Extending the range

In 1997, Awschalom's team reported that electron spins could survive in a coherent state for more than 100 nanoseconds⁷. Then last year Awschalom and his colleague James Kikkawa managed to transport a coherent 'spin packet' of electrons over a distance comparable to the size of a typical electronic device⁸. They used a laser pulse to excite a spin packet in a sample of GaAs, and dragged it along using an electric field.

Awschalom's team has also succeeded in trapping coherent spin packets in semiconductor 'quantum dots' — crystals of cadmium selenide between 2 and 8 nanometres across⁹. This shows that qubits could be temporarily stored in a quantum computer without becoming scrambled.

Awschalom believes that creating coherent spin packets using light and probing for their existence optically, as his team has done, will become increasingly important. It will allow spintronics to interact with the field of optoelectronics, important in

Spin is an ideal candidate for the quantum bits of quantum computing.

telecommunications. Information is generally transmitted over long distances as a series of light pulses in optical fibres. But at present, the signals must be converted back to electronic form for processing and storage, and this is an inefficient process. "The bottleneck is the photon-to-electron conversion," Awschalom explains. He believes this conversion can be achieved more efficiently in optically controlled spintronic devices.

For quantum computing, meanwhile, many of the theoretical proposals invoke the manipulation of the spins of atomic nuclei, rather than those of electrons. This can be done using nuclear magnetic resonance, which usually relies on flipping nuclear spins using pulses of radio waves. However, Awschalom and Kikkawa have shown that it can also be achieved using light: by pulsing an optical beam at an appropriate resonant frequency, they were able to flip nuclear spins in GaAs¹⁰. The light pulses create packets of spin-coherent electrons, whose magnetic fields then influence the nuclear spins.

While practical quantum computers remain a distant prospect, spin coherence could be applied in simpler spintronic devices. Awschalom foresees a "genuine quantum electronics". Rather than treating electrons like tiny billiard balls, this would exploit their wave-like properties. In a coherent spin packet, the electron waves are all in phase. This might offer a new switching mechanism, using destructive interference effects to 'cancel out' two out-of-phase coherent spin packets. Awschalom predicts that this will offer "completely new kinds of functionality" in microelectronic and optoelectronic devices. For example, a team led by Sasha Hallstein of the Max Planck Institute for Solid State Physics in Stuttgart has already used electron spin coherence to construct an ultrafast optical switch¹¹.

With GMR read-out heads now in widespread use and MRAM devices soon to reach the market, spintronics is already starting to make its presence felt. Whether the microelectronics industry will embrace the vision of quantum electronics remains to be seen. But Wolf at DARPA is confident that the future is bright. "Spin hopefully will provide a host of new devices for very-high-speed, low-power electronics and optoelectronics," he says.

Philip Ball is a Consultant Editor of *Nature*.

1. Cowburn, R. P. & Welland, M. E. *Science* **287**, 1466–1468 (2000).
2. Ohno, H. *et al. Appl. Phys. Lett.* **69**, 363–365 (1996).
3. Akiba, N. *et al. J. Appl. Phys.* (in the press).
4. Hammar, P. R., Bennett, B. R., Yang, M. J. & Johnson, M. *Phys. Rev. Lett.* **83**, 203–206 (1999).
5. Fiederling, R. *et al. Nature* **402**, 787–790 (1999).
6. Ohno, Y. *et al. Nature* **402**, 790–792 (1999).
7. Kikkawa, J. M., Awschalom, D. D., Smorchkova, I. P. & Samarth, N. *Science* **277**, 1284–1287 (1997).
8. Kikkawa, J. M. & Awschalom, D. D. *Nature* **397**, 139–141 (1999).
9. Gupta, J. A., Awschalom, D. D., Peng, X. & Alivisatos, A. P. *Phys. Rev. B* **59**, R10421–R10424 (1999).
10. Kikkawa, J. M. & Awschalom, D. D. *Science* **287**, 473–476 (2000).
11. Hallstein, S. *et al. Phys. Rev. B* **56**, R7076–R7079 (1997).

How diagnosis with microarrays can help cancer patients

Sir — Alizadeh *et al.*¹ show how the analysis of gene expression using high-density DNA microarrays can improve the diagnostic accuracy of diffuse large B-cell lymphoma. But there are many problems to be overcome before this approach may become generally applicable to cancer diagnosis, as hoped for by Berns in his interesting News and Views article "Gene expression in diagnosis"².

Analysis of gene expression using DNA microarrays is unlikely to replace histopathology as the prime indicator of prognosis. In relation to cancer, the histopathologist makes three important judgements: diagnosing the cell of origin (tumour type) of the cancer; how closely the tumour resembles the tissue of origin (tumour grade); and the extent or site to which the cancer has spread (tumour stage). The extent or stage of the cancer is by far the most important judgement in most types of cancer; it dictates treatment and is an accurate predictor of the prognosis.

Tumour grade also correlates strongly with prognosis, but this is a subjective assessment, which limits its reproducibility and clinical value. It is in grading tumours that microarrays are most likely to improve the accuracy of prognosis.

It is not by chance that the first applications of microarrays to the diagnosis of cancer were made on leukaemias and lymphomas^{1,3}. These cancers tend to be single cells that can be obtained non-invasively in a blood sample and can be separated to high purity using cell-surface markers. However, 90 per cent of cancers are solid tumours that have to be surgically removed and are extremely difficult to purify. The microarray approach can provide only a crude average of gene expression across all the cells used to prepare the RNA or DNA. Unless the cancer-cell preparation is highly pure, the contribution of the myriad other cells within the cancer (normal cells, supporting stroma, blood vessels, lymphocytes, and so on) can mask the expression pattern on the array. Pure cancer cells can be obtained by laser microdissection of tissue sections or by cell sorting, but these are labour-intensive and time-consuming processes.

Another problem is the variability in invasive potential between cancer cells within the same tumour. For example, prostate cancer is often present in many distinct foci, only one of which may have the potential to be invasive and dictate the outcome for the patient. Must each focus be

analysed separately? Within one focus there may only be a few cells with the potential to invade — will their gene-expression pattern be masked by the surrounding, less malignant cells?

There are already a vast number of prognostic markers available for every type of cancer, and many of these are of independent prognostic significance. However, they are of little use to the individual patient because, although statistically significant, they do not provide the quality of information needed to be confident that a major operation, for example radical prostatectomy or cystectomy, is beneficial.

The acid test for DNA microarrays in cancer diagnosis will be whether the information they provide alters the patient's treatment — a likely outcome but yet to be proven. Using pure populations of cancer cells (for example, cancer-cell lines, flow- or magnetically sorted cancer cells, or cells obtained by microdissection), new and more powerful prognostic markers will be identified using microarrays, as demonstrated by Alizadeh *et al.*¹. It will then be possible to use conventional methods on tissue sections (such as immunohistochemistry and *in situ* hybridization) to apply these new markers to individual cancers.

John R. W. Masters*, Sunil R. Lakhani†

**Institute of Urology and Nephrology, University College London, 67 Riding House Street, London W1P 7PN, UK*

†*Department of Histopathology, University College London, Rockefeller Building, University Street, London WC1E 6JJ, UK*

1. Alizadeh, A. A. *et al. Nature* **403**, 503–511 (2000).

2. Berns, A. *Nature* **403**, 491–492 (2000).

3. Golub, T. R. *et al. Science* **286**, 531–537 (1999).

Short-sighted move to close the 12-m telescope

Sir — Recent budgetary choices have forced the closure of the US National Radio Astronomy Observatory 12-m telescope in Kitt Peak, Arizona. As this detrimentally affects astronomy in many ways, I am writing to oppose this move.

The 12-m telescope remains one of the most important and vital components of US millimetre astronomy. For graduate students in the United States, it is the only accessible, competitive, peer-reviewed facility capable of both training them as future scientists and providing for internationally recognized research. The 12-m telescope is a vital component to my own PhD thesis and much of my future work is predicated on its existence. Specifically, wide-field mapping and full synthesis imaging will not be possible without it.

Alternatives such as CSO, JCMT, HHT

and IRAM are inadequate. None of them except IRAM 30-m has the frequency coverage of the 12-m. For example, deuterium-bearing molecules are often found to have ground-state transitions below 85 GHz, the limiting tuning range of many other receiver systems. The deuterium component of the interstellar medium (ISM) — as a measure of primordial nucleosynthetic products, chemical evolution in the ISM, and observational limit on abundance predictions from star-formation theories — will no longer be accessible.

The alternative facilities noted above are less accessible to US astronomers than the 12-m telescope, either because the observatories have closed associations with their parent institutions or because they are very expensive to reach. Furthermore, the HHT, though a capable enough system, lacks receivers at 2-mm and 3-mm wavelengths to compensate for the loss of the 12-m telescope.

Two solutions are apparent. The first is to establish a consortium of universities to take over operation of the 12-m. Some emergency funding, even if outside the National Science Foundation's budget, must be obtained in the interim period in order to allow minimal operation of the 12-m. This will also allow NRAO to retain the staff, engineers and scientists whose vast millimetre knowledge will disappear if the closure goes through as planned.

Many of us have written to counteract the impression that the 12-m telescope is no longer a vibrant instrument, or that NRAO has overextended itself. The national observatories maintain widely accessible, strong facilities with flexibility and extensibility. The 12-m telescope remains important until work begins of the construction of the Atacama Large Millimetre Array (ALMA) — 64 antennas located at 5,000 m in Chile's Atacama Desert.

Ronak Shah

Department of Astronomy, University of Virginia, 530 McCormick Road, Charlottesville, Virginia 22903, USA

Genetic modification and the meat market

Sir — Jonathan Latham¹ in Correspondence states that famine is a problem of global food distribution and arable efficiency rather than of food quantity, and that hence there is no need for genetically modified (GM) crops. Although this argument may in theory apply today, it will not in practice apply tomorrow.

Of course there are scientific questions that still need to be answered about GM technology, but we already know that by

▶ the middle of the twenty-first century the world is going to face a food crisis, and that agriculture will consequently put increased pressure on wildlife habitats.

In 1998, the UK Institute of Biology and six affiliated societies (whose specialist interests range from agricultural production to ecological conservation) produced a report on the social and ethical aspects of GM crops². We cited half a dozen indicators of the forthcoming shortfall in global food supply, including the following. Forty per cent of terrestrial primary productivity is already managed by humanity. The trend for the past 15 years has been a reduction in grain production per capita. Global sea-fish catches have been in steady decline since 1990 because of over-fishing. World carry-over stocks of grain are declining from one year to the next. The grain harvest area per person has been declining since the late 1970s, owing to increasing population, growth in industry and desertification.

The increasing consumption of meat in the rich nations has put more pressure on the poor, although reversing this trend alone (even if it were realistic) would not counter the pressures caused by a population increase of 40 to 80 per cent over the next four decades. The world shows no sign of turning vegetarian. Although I am sympathetic to Latham's conclusion that "what is missing is the 'purchasing power' of the poor", the evidence is that when the poor become a little richer they eat more meat.

Given that agricultural inefficiencies and global inequalities are bound, sadly, to continue, it is likely that genetic modification where appropriate will make a significant contribution to human well-being — and to that of other species.

Jonathan Cowie

*Institute of Biology, 20–22 Queensberry Place,
London SW7 2DZ, UK*

1. *Nature* **404**, 222 (2000).

2. *GM Crops: The Social and Ethical Issues* (Institute of Biology, 1998). www.iob.org/gmcrops.html

Distinguished scientists back Germany's DFG...

Sir — Your recent News report "German research agency stifles creativity" (*Nature* **404**, 217; 2000) gives a negative and incorrect impression of the Deutsche Forschungsgemeinschaft (DFG).

Nature claims that DFG's inability to assess novel research areas and interdisciplinary research areas threatens career opportunities, especially for young researchers. The cases mentioned in the *Nature* report, however, are neither representative nor described in an unbiased manner.

Typically, the reviewing process of the DFG takes less than six months and involves a large number of scientists from foreign and German institutions and from senior as well as junior ranks. Every attempt is made to support the best and the most innovative scientific proposals. In fact, time and again high-risk proposals are funded that, for example, would have no better chance of support from the US National Institutes of Health.

Of course, no system is free of errors, and occasional undeserved negative judgements may be made. However, continual efforts are made to improve the system. Overall, we are impressed by the flexibility of the DFG, its unbiased support for creative, high-quality research and its programmes for young scientists and interdisciplinary research even at times when its budget is tight.

At this juncture, our most urgent concern is to convince politicians to increase funding to the DFG significantly. This is particularly important for the support of young scientists. We are very proud of the DFG as a self-governing body of the German scientific community and we believe it to be, by any standards, one of the best scientific funding agencies.

Reinhard Jahn

*Max Planck Institute for Biophysical
Chemistry, Am Fassberg, D-37077
Göttingen, Germany
rjahn@gwdg.de*

Other signatories of this letter:

August Böck Ludwig-Maximilians-Universität München

Klaus M. Breiner Emmy-Noether-Fellow, ETH Zürich

Karin D. Breunig Universität Halle-Wittenberg

Herman Bujard Universität Heidelberg

José Campos-Ortega Universität zu Köln

Detlev Ganten Max-Delbrück-Zentrum, Berlin-Buch

Ingrid Grummt Deutsches Krebsforschungszentrum, Heidelberg

Peter Gruss Max-Planck-Institut für biophysikalische Chemie,
Göttingen

Christo Goridis CNRS, Marseille

Robert Huber Max-Planck-Institut für Biochemie, München

Michael Hoch Gerhard-Hess-Fellow, Universität Bonn

Herbert Jäckle Max-Planck-Institut für biophysikalische Chemie,
Göttingen

Regine Kahmann Ludwig-Maximilians-Universität
München

Claudia Koch-Brandt Universität Mainz

Maria Leptin Universität zu Köln

Hartmut Michel Max-Planck-Institut für Biophysik, Frankfurt

Angelika Noegel Universität zu Köln

Erwin Neher Max-Planck-Institut für biophysikalische Chemie,
Göttingen

Christiane Nüsslein-Volhard Max-Planck-Institut für
Entwicklungsbiologie, Tübingen

Nikolaus Pfanner Universität Freiburg

Konrad Sandhoff Universität Bonn

Astrid Schoen Heisenberg fellow, Universität Würzburg

Petra Schwill Biofuture junior group, Göttingen

Kai Simons Max-Planck-Institut für molekulare Zellbiologie und
Genetik, Dresden

Eberhart Zrenner Universität Tübingen

Signed on behalf of 1,164 other biomedical scientists. The full list of
names is available from R. J.

...but young researchers feel disillusioned

Sir — Your recent News report "German research agency stifles creativity" (*Nature* **404**, 217; 2000) gives a negative impression of the Deutsche Forschungsgemeinschaft (DFG) — but one that is, in our experience at least, correct.

Nature claims that the process threatens young researchers' career opportunities in particular. Our last four applications for grants in the area of environmental toxicology (mechanisms of microcystin toxicity in the aquatic environment) were rejected, after an average delay of 10–12 months, as "irrelevant" or "dealing with non-existent problems". We did, fortunately, receive support for a similar grant from the European Union; the results of these studies have been or will be published this year, and they form the basis of an EU patent application.

The referees of our unsuccessful DFG applications did not seem, to us, to be up-to-date in their knowledge of the topic, or they had little understanding of environmental toxicology. Indeed, the comments we received from the DFG made us wonder whether the referees had even read the grant. They were so contradictory of each other as to provide us with no constructive advice on how to improve the application. The upshot was that, while we were able to demonstrate that our proposed research could be done, and was publishable in peer-reviewed journals, it was not considered fundable by the DFG. This kind of outcome may not seem devastating to seasoned scientists with established careers. But it impedes the careers of young researchers dependent on DFG funding within Germany, and is demotivating.

A better approach would be for grants to be sent out for review internationally; for referees' comments to be sent to the applicants in their original form, not rewritten by DFG to maintain anonymity (we are happy for peer-review to remain anonymous, but the rewriting leads to incomprehensible comments); and, as proposed in the *Nature* report, for applicants to be able to attend referees' meetings to answer questions and defend their grants.

Daniel R. Dietrich, Bettina C. Hitzfeld

*Department of Environmental Toxicology,
University of Konstanz, PO Box 5560 -X918,
D-78457 Konstanz, Germany*

Nature replies — The *Nature* report states explicitly that the DFG reviewing process averages five to six months. The complaints discussed in the article concern the outliers to this average — applications in new, interdisciplinary, not traditional, areas of research. ■

'Marketing' species conservation

Financial incentives can be found to conserve a species threatened by trade.

**R. E. Gullison, R. E. Rice
and A. G. Blundell**

The difficulties of reconciling international trade with its environmental impacts were memorably highlighted recently by street demonstrators at the World Trade Organization meeting in Seattle. This outcry, however, was only the latest in a long line of attempts to protect at-risk species and ecosystems from the ravages of the market. As delegates from around the world leave this month's meeting of the Convention on International Trade in Endangered Species (CITES) in Nairobi, has any progress been made in resolving this dilemma?

Unfortunately, although regulatory protections were increased for some important species, for many others the issue remains unresolved. In fact, despite consumer boycotts and international agreements such as CITES, most efforts to promote more sustainable use of natural resources have failed, often for the same reason — they have not provided direct incentives for conservation. However, recent examples of conservation agreements and land purchases are concrete evidence that direct financing can be a viable and cost-effective approach to conservation.

The case of mahogany

The New World mahoganies (*Swietenia* spp.) provide a telling example of the difficulties encountered in conserving commercially important species. Mahogany, a highly valued hardwood, has been traded in the western world for more than 500 years. However, throughout this period, trade has been maintained not by sustainable production but by



Just one commercially viable species now remains to maintain the 500-year-old mahogany trade.

continuous local depletion and shifting sources of supply¹. As early as 1735, mahogany was becoming scarce in Jamaica and operations were shifting to Central America. The Caribbean species (*S. humilis* Zucc.) and Pacific coast species (*S. mahagoni* L. Jacq.) are now commercially extinct; nearly all current production is from a third species (*S. macrophylla* King) in South America².

Much time and effort have been devoted to promoting the adoption of more sustainable management practices, with some effect in temperate forests. But loggers in the tropics have generally chosen not to adopt sustainable management³ because, with limited financial incentive and government oversight, conventional logging is more attractive. Sustainable-management policies with

their harvest restrictions are more costly and hence less profitable to local industry than is conventional logging⁴. Government regulation and campaigns by non-governmental organizations may ultimately achieve a situation where sustainable logging is more widely adopted in the tropics, but this is unlikely to be in time to help mahogany and many other species at risk.

Although CITES has been quite effective for some species, such as elephants and whales, it has done disappointingly little to help the conservation of mahogany and other timber species. Despite a rapidly shrinking commercial range, poor regeneration following harvest, and an almost complete absence of control of logging, efforts to list mahogany as a protected species on CITES (Appendix II) have failed — not once but three times.

Objections to the listing have ranged from claims that international trade does not threaten the species, to the view that CITES should not or cannot control the timber trade, or even that trade restrictions might prove counterproductive. A more significant obstacle, though, is the conflict of interest that arises when a country votes on conservation measures that directly affect trade in a species from which it benefits financially. Whatever the causes, the failures at CITES and the pressure for profit from mahogany have been so effective a deterrent to conservation that listing mahogany was not even on the agenda at Nairobi.

The ineffectiveness of international agreements has led some environmentalists to encourage boycotts. In the early 1990s, for example, the "Mahogany is Murder" campaign led by the UK Friends of the Earth



The pressure for profit from mahogany has significantly undermined conservation efforts.

focused world attention on unsustainable logging and deforestation, and over a five-year period, UK imports of mahogany were reduced by 95% (ref. 5). Although enormously successful in that one country, the effect of the boycott has now been almost completely offset by increased consumption in the United States. For boycotts to work, they must affect the majority of importing countries, with few or no alternative markets available. In the case of mahogany, even a global boycott would not solve the problem because it would provide no direct incentive for mahogany conservation. The species would still be subject to harm from logging for other species, and to pressures to convert forests to other uses.

Financial incentives

The lesson from the shortfalls of CITES, boycotts and sustainable management in protecting mahogany is that conservation must be the direct objective of policy, rather than a desired by-product, to ensure that trade does not threaten valuable species. Even more important, policies must provide a positive inducement to drive the widespread adoption of conservation. Existing initiatives now require that resource owners conserve for no reward, or even at a net cost, but only tangible incentives for conservation will induce local communities, organizations or countries to implement concrete efforts to conserve species of economic importance.

How might such a strategy be implemented? Perhaps the simplest, most effective approach is to finance a 'safety net' of protected populations. Two recent examples from Bolivia demonstrate that feasible options exist for affordable ways to finance the protection of mahogany.

First, in 1998, The Nature Conservancy, an international conservancy group, doubled the size of Bolivia's Noel Kempff Mercado National Park by acquiring the logging rights to 630,000 hectares of adjacent timber concessions for only US\$2.50 per ha. Here, isolated forests containing remnant adult populations and stands of pre-commercial mahogany trees were bought very cheaply. Second, Conservation International last year bought the logging rights to a 45,000-ha concession that was later added to Madidi National Park for \$2.22 per ha. Both transactions gave the owners concrete financial incentives and resulted in the protection not only of mahogany but also of countless other species.

Conservationists are pioneering other mechanisms for funding the protection of endangered species on community and indigenous lands. One model is the recent landmark arrangement between six environmental organizations and a small community in Mexico by which the Cebadillas community will receive \$250,000 over 15 years to preserve the nesting habitat of the western thick-billed parrot (*Rhynchopsitta pachy-*



For mahogany, a global boycott on trade would provide no direct incentive for its conservation.

rhyncha). The area, 2,400 ha of old-growth forest in Northern Chihuahua, comprises about 50% of the remaining nesting habitat for the parrot and will help ensure the species' long-term survival.

This innovative agreement shows that landowners can be given a direct incentive to protect their forests when they obtain sufficient income from conservation. One of Brazil's largest remaining intact populations of mahogany, in the Kayapo indigenous territory, is appropriate for this kind of incentive.

Money for conservation

We believe that such initiatives can be easily financed. The public is very willing to fund environmental protection: in 1999, for example, The Nature Conservancy generated \$700 million to acquire and protect habitats in the United States and elsewhere. In January, Broward County in Florida proposed spending \$400 million to buy 1,270 ha of natural habitats and farmlands at roughly \$157,000 per ha. These examples suggest that plentiful funding should be available for conserving species such as mahogany that have a high international profile. Further, as suggested by the Latin American examples, conservation can cost much less than is commonly believed.

Direct intervention is not likely to be welcomed by everyone. Foreign funding could be seen as a form of neocolonialism, and some would argue that the most appropriate role of conservationists is to persuade the market to accept timber only from sustainably managed forests. However, in our view, paying for direct protection is simply a practical means of compensating those who conserve resources of global importance. Directly financing protection, moreover, can reinforce alternative approaches to conservation such as sustainable management, by providing a powerful hedge against their failure. If boycotts, sustainable use and CITES continue to be ineffective for mahogany, the future of the species can still be secured by protecting viable populations in nature reserves.

In this way, direct financing of protection

follows the 'precautionary principle', an environmental-protection strategy that is receiving much backing in international legal agreements. The precautionary principle requires that prudent measures be taken to avoid the risk of irreversible damage such as the loss of a species. In the case of mahogany, a safety net of protected populations would greatly reduce the risk of irreversible damage from continued exploitation.

We feel there is an urgent need to reconsider how CITES and other mechanisms work to ensure the conservation of commercial species. To take mahogany as an example, although the signatories to CITES have again failed to list the species on Appendix II, a group of technical experts is to study its situation more closely. We suggest that this working group focuses on identifying remaining intact populations and determining the minimum critical size of viable populations that should be protected.

CITES cannot be effective alone, however. Local and national governments, conservation organizations and the private sector must also help to transfer funds from those who value species such as mahogany to those who are best able to conserve representative populations but who need incentives to do so. ■

R. E. Gullison is at the Centre for Biodiversity, University of British Columbia, 6270 University Boulevard, Vancouver, British Columbia V6T 1Z4, Canada. (e-mail: gullison@island.net). R. E. Rice is at the Center for Applied Biodiversity Science, Conservation International, 2501 M Street NW, Washington, DC 20037, USA. A. G. Blundell is at the National Center for Environmental Assessment, US Environmental Protection Agency, 401 M Street SW, Washington, DC 20460, USA.

1. Lamb, F. B. *Mahogany of Tropical America: Its Ecology and Management* (Univ. Michigan Press, Ann Arbor, 1966).
2. Rodan, B. R., Newton, A. C. & Verissimo, A. *Environ. Conserv.* **19**, 331–342 (1992).
3. Bowles, I. A., Rice, R. E., Mittermeier, R. A. & da Fonseca, G. A. B. *Science* **280**, 1899–1900 (1998).
4. Pearce, D., Putz, F. & Vancley, J. K. *A Sustainable Forest Future* (CSERGE Working Pap. GEC 99-15, Univ. College London, 2000).
5. Schwartzman, S. & Kingston, M. *Global Deforestation, Timber, and the Struggle for Sustainability* (Environmental Defense Fund, New York, 1997).

Sir James in cyberspace

Let there be public revelry — the *OED* takes to another dimension.

Oxford English Dictionary Online

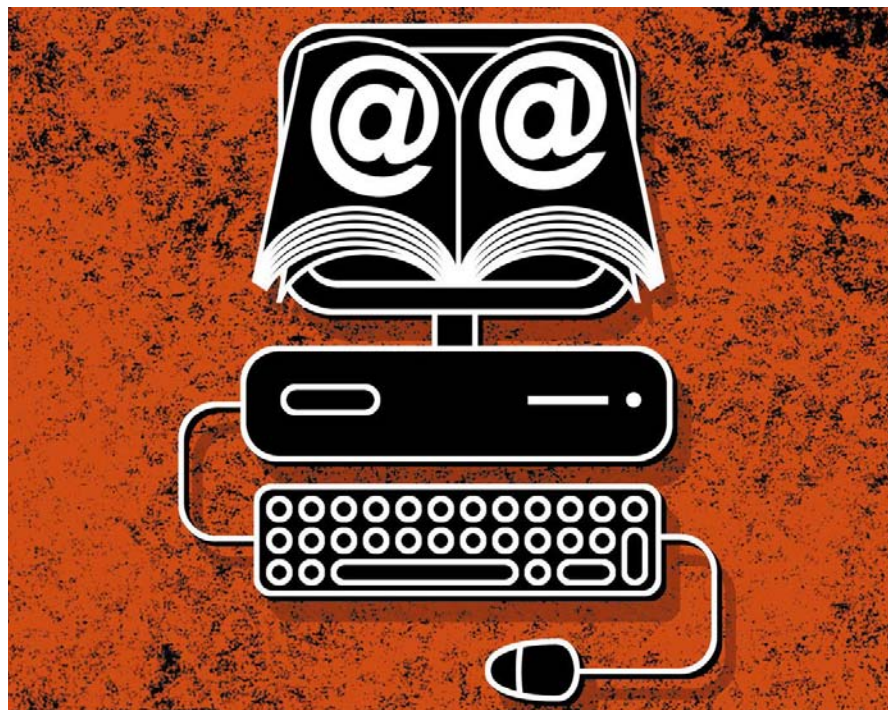
Oxford University Press: 2000. Annual subscription £350+VAT, \$550

Walter Gratzner

It was said of Thomas Macaulay that he not only overflowed with learning but stood in the slop. This is precisely what happens when you first uncork the new online-only edition of the *Oxford English Dictionary* (*OED*), and allow it to spill a measure of its profligate riches over your feet.

The *OED*, let it be said, is an incomparable monument of scholarship, one of the wonders of our age. I began my exploration of its vast territories by checking some cherished etymological beliefs, or, as I now find, misapprehensions. Look up *loo* for instance and you find that it is first of all a card game; the description is detailed enough for you to be able to sit down to a hand or two with the characters in works by William Wycherley or Jane Austen, who cavort in the quotations. Then again, a *loo* (*Obs. exc. Hist.*) was “a velvet mask partly covering the face, worn by females in the 17th century to protect the complexion”. The third *loo* is “the name given in Bihar and the Punjab to a hot dust-laden wind”, and to prove it, there is a stirring passage from Kipling. And finally, a *loo* is indeed “a privy, a lavatory”. Here the *OED* is sparing of etymologies, but vouchsafes that A. S. C. Ross examined possible sources in 1974 in the October issue of *Blackwood's Magazine*, wherein he concludes that it derived “in some manner that cannot be demonstrated from *Waterloo*”; no *gardey loo*, then (the cry of the Edinburgh housewives as they emptied their chamber pots into the street from the upper window)? Or even *l'OO*, from the two holes in a French privy door that allowed you to see whether it was occupied? Alas, it seems not.

Sidling towards science, I wandered at random among the units of measurement — pascals, oersteds, newtons, Bohr magnetons, units SI and units obsolete — the *OED* embraces them all. But for a more stringent test I sought out the eccentric *Dent Dictionary of Measurement* and found first the unit of pain, the dol: this the *OED* effortlessly defined. Moving a little further from the beaten track, I challenged it with *cran*, the unit for the measurement of herrings, and was rewarded by the following definition: “A measure of capacity of fresh herrings as caught; fixed by the Fisheries Board at 37½ gallons (about 750 fish)”. “Up to 1815,” the dissertation continues, “the *cran* was measured by heaping full a herring-barrel with the ends taken out, which was then lifted,



leaving the heap on the ground or floor. In 1816, the Commissioners for the Herring Fishery fixed the capacity of the ‘cran’ at 42 gallons, Old Wine Measure, which in 1832 was raised to 45 gallons, 42 gallons when ‘pined’ being found insufficient to make a barrel of bung-packed herrings.” There is more, not forgetting the quotations. A *cran* is also a term (*Obs.*) for the crane and the heron. On the other hand, there is no mention of the *cran*, defined by *Hobson-Jobson*, the ebullient nineteenth-century dictionary of Indian and Malaysian words, as “A modern Persian silver coin”. The *Dent Dictionary* also offers a list of collective nouns: wildfowl travel in *sutes*, turtles in *dules* and jellyfish in *smucks*. The *OED* lists none of these, but it concedes that a brood of ducks was at one time a *badelyng*, and of pheasants a *nye*.

Did the editors of *Dent* perhaps make up the others? Or could they all be usages so far decayed as no longer to count as words? For the *OED* is a living organism, which not only grows, but also presumably sheds detritus. The enterprise (originally *The New English Dictionary on Historical Principles*) was conceived in this sense by a group of Victorian scholars in 1857; their plan was to engage volunteers, allocating to each one a sector of the English literary and historical landscape in which to forage.

In 1878 the Oxford University Press was finally persuaded that the project had merit, and chose the redoubtable schoolmaster James Murray as editor. The task was not

completed until 1928, long after Murray’s death. It was never the founding fathers’ intent to lay down rules of linguistic usage. The dictionary was to be authoritative, but not authoritarian, to observe and describe, not prescribe. Ambrose Bierce, in *The Devil’s Dictionary* (Wordsworth), defined a dictionary thus: “A malevolent literary device for cramping the growth of a language and making it hard and inelastic”.

The *OED* has been the very antithesis. It abjured from the outset any desire to emulate the French Academy, with its Forty Immortals, who in the eighteenth century saddled their language with the curse of the circumflex and even in our time were prepared to instruct scientists that an enzyme was feminine, when all French biochemists knew perfectly well that it was *un enzyme*. The *OED* does not spurn those neologisms that so affront the guardians of linguistic purity. Thus, as one of the definitions of *hopefully* it offers: “It is hoped (that); let us hope... orig. U.S. (Avoided by many authors)”; it cautiously observes.

It is clear that the organism has been almost continuously updated, with the various additions, which followed the majestic second edition of 1985 in 12 volumes, with CD-ROM as an option. The new online dictionary is still undergoing revision, a labour that will take another 10 years to complete. So far, only M–MAH is finished, and the editors concede that the *OED*’s content of science may still be found in some degree wanting.

Look up *scientist*, then, and you will

discover the history of the word from its first tentative appearance in 1834: an article in the *Quarterly Review* laments the fragmentation of science, as reflected in “the want of any name by which we can designate students of the knowledge of the natural world collectively... some ingenious gentleman proposes that, by analogy with *artist*, they might form *scientist* ... but this was not generally palatable”. In 1840 William Whewell (said to be the last man to know everything) re-invented the word and in time it stuck. But look down the column of adjoining entries on your screen and you can find what we were spared: “*Sciencer, Obs.* A professor of a particular science”. And then there is the magnificent, and much-needed, conceit, “*Scientaster* [... after *poetaster*] A petty or inferior scientist”, coined by the physiologist Michael Foster in 1899; a quotation follows from Foster’s biography of Claude Bernard.

The physical sciences seem to me to do better on the whole than biology, perhaps because the corpus of knowledge is less diffuse. Here, under *flavour*, are the six flavours of quarks; *string theory* and *superstrings* get crisp definitions; *Higgs* appears, attached to *boson*, *field*, *mechanism* and *particle*; and, going back a little in time, *Debye* surfaces under *Debye effect*, *Debye–Hückel theory*, *Debye–Scherrer method*, *Debye temperature* and *Debye unit*. Here, too, is *buckminsterfullerene*, complete with a quotation from Harry Kroto in *Nature*. The besetting fatuity, still found in many dictionaries, of recording stoichiometric rather than structural chemical formulae has been largely but not entirely expunged, so the peptide *melittin* from bee venom comes out as $C_{131}H_{229}N_{35}O_{31}$ (who counted?). *Sulfur* — in the American and now internationally sanctioned spelling — is not in evidence. Prefixes for large units do better than those for small: we find *giga-* and *tera-*, but not *atto-* and *zepto-*.

As to biochemistry, all the most familiar proteins seem to be present and quite a number of others, although some definitions (*actin* for instance) are in sore need of revision. We have *transferrin*, *ferritin*, *laminin*, *reverse transcriptase*, *spectrin* and *ankyrin* even, but no *integrin*, *fibronectin*, *clathrin*, *calpain* or *G-protein*. *Oncogene* and *homeobox* are in, but *apoptosis* is missing and so are *T-cells* or *T-lymphocytes*, *genomics* and indeed *PCR*. In the revised M–MAH sliver, I looked for and found *magainin*, first isolated from the skin of the “clawed toad” — but *Xenopus*, I am assured, is not a toad, but rather a clawed-toed frog. Missing is *MHC* (major histocompatibility complex), ubiquitous enough, arguably, to qualify for inclusion. Of course, the *OED* does not purport to be a textbook or encyclopaedia of science and somewhere must draw an arbitrary line between the barely useful and

the totally recondite, but among the lacunae are expressions that a journalist, for instance, might well want to track down.

The dictionary is diverting on misuses that have become irretrievably embedded in common speech. *Parameter* (first spotted in a mathematical tract in Latin by one C. Mydorge in 1631) receives separate definitions in conic sections, crystallography, mathematics, electricity and statistics, but also “In extended use: any distinguishing or defining characteristic or feature ...”, with a quotation, fittingly enough, from *New Sociology* (“We would then say that a social theory has a human-nature parameter” — ah, so!). A *quantum jump* is not only a transition between stationary states of a quantized system, but more especially “*transf.*, a sudden large increase or advance”, also now known to politicians and estate agents as a *quantum leap*.

The online format of the *OED* is friendly and responsive. A click of the mouse will bring up or hide pronunciation, etymology, quotations (2.5 million of them and, to my mind, the greatest treasure of all) and a date chart showing the development and decline of usages. You may retrieve quotations from any one author, and relate them to particular words; you may bookmark entries and you may search for your favourite cliché (“sick as a ...”). If you are unsure of a spelling you can enter a question mark in the middle of your word or use an asterisk (wild card) to denote an undefined number of letters. This will allow you, if you are so inclined, to play word-games. So, for illustration, a reader in a newspaper recently asked whether any words existed in which a single consonant appeared three times in tandem. Well, a brief search of the *OED* for words of the type, ‘sss’ at once yielded *bossship* and one typo.

Today’s science, with its headlong pace of progress and its ever-shifting frontiers, probably makes impossible demands of the *OED*’s editors, and it will be interesting to see how they grapple with it between now and 2010.

Meanwhile, we should celebrate a great and noble assertion of intellectual virtues and an inexhaustible source of pleasure, to those at least who can afford to or otherwise get at it. H. L. Mencken, journalist, lexicographer and sage, thought that the completion of the first edition of the *OED* in 1928 should be marked in Oxford by public revelry — “military exercises, boxing matches between the dons, orations in Latin, Greek, English and the Oxford dialect, yelling matches between the different colleges, and a series of mediaeval drinking bouts”. I fancy I can see the benign shade of Sir James Murray, surrounded by his team of indexers, celebrating out in cyberspace with a small dry sherry. ■

Walter Gratzer is at the Randall Institute, King’s College London, 26–29 Drury Lane, London WC2 5RL, UK.



When silence is not a true option

The Woman Who Knew Too Much: Alice Stewart and the Secrets of Radiation

by Gayle Greene

University of Michigan Press: 2000. 336 pp.

\$35, £19.95

Sunetra Gupta

“Practising medicine without asking these larger questions is like selling groceries across the counter,” said Alice Stewart when asked why, at the peak of a career in clinical medicine, she had decided to abandon it in favour of practising epidemiology.

In 1945, the importance of identifying the risk factors of infectious diseases was becoming obvious, and efforts had already been made to define and understand the ecological processes underlying the spread of infection. The Institute for Social Medicine had been established at the University of Oxford in 1943, reflecting the recognition that it might also be worthwhile to investigate the causes of non-infectious diseases such as cancer. These diseases might have “discoverable origins in social, domestic, or industrial maladjustment”, according to the institute’s founder, John Ryle. Ryle died in 1950, and the institute was diminished to the Social Medicine Unit and its building taken away. Stewart, who had been Ryle’s assistant, was given a readership and made its head with a budget so small that there was “barely enough to light a gas fire”.

It would have been perfectly possible for Stewart at this time to keep up some semblance of research and devote the rest of her time to her country garden, not to mention the lively intellectual circuit in which she had a singular place as the lover of the distinguished literary critic, poet and mathematician William Empson. However, according to her biographer, “epidemiological investigation engaged her like a piece of detective work”, and in 1950 Stewart set about organizing a retrospective case control study to identify risk factors for childhood cancer on a grant of £1,000 from the Lady Tata Memorial Fund for Leukaemia Research.

“I spent those £1,000 on railway fares traveling the length and breadth of England, going to each public health official, saying ‘here are the questionnaires, will you help?’,” said Stewart. From this incredible effort came the startling revelation that a single obstetric exposure to X-rays significantly increases the risk of an early cancer death. The Oxford Survey of Childhood Cancer, as it came to be known, continued for 30 years, beyond Stewart’s retirement in 1974.

She relocated to the University of Birmingham, and found “an empty corridor

[illegible]

progressive public school, and came to Cambridge already integrated into a “charmed inner circle ... of beautiful well groomed young women”. She married a fellow student at Cambridge who became a French master at Harrow School, and for a while she played the role of wife and mother. In the meantime, she hoisted herself out of a low-key research job at the London School of Tropical Medicine and Hygiene into full-time clinical medicine. She sat for her membership of the Royal College of Physicians and succeeded in her first attempt, and in 1939 was appointed to a consultant post.

The quality of her life before the war and during her years in Oxford does little to substantiate the suggestion of persecution in the title of the book. It is only in the latter part of the book that we begin to identify with the sinister ring of the title, when Greene lays bare the extraordinary history of resistance by the US government to the progress of research on the hazards of radiation. Stewart's place in this scenario was as a proponent of the notion that sustained exposure to low-level radiation carries an increased risk of cancer. She became vocal in this post-retirement period of her life in supporting the various researchers who had been isolated by this process of "governmental interference". She began to serve regularly as an expert witness in compensation cases, and engaged at an international level in the public debate about nuclear safety.

Greene attempts to comment on this

Sunetra Gupta is at the Wellcome Centre for Infectious Disease Epidemiology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

by Abraham Pais

Oxford University Press: 2000. 364 pp.
£25, \$30

Having paid compulsory tribute to the Medicis, the Grand Dukes of Tuscany, Giorgio Vasari dedicates his monumental opus, *Lives of the Most Excellent Painters, Sculptors and Architects* (Giunti, Florence, 1568), to his fellow artists. Vasari was a great biographer of the artists of the Italian Renaissance. In 1,012 quarto pages, with illustrations, he turned their life stories into gripping tales. He could do this because he was one of them, a famous artist in his own right, for whom Michelangelo drew fresco cartoons. He was also a magnificent writer.

Not all biographers had such a style. Plutarch and Tacitus, who wrote biographies of the great men of the first century, were not emperors themselves, although they obviously moved in the right circles — Tacitus, for example, was the son-in-law of Emperor Agricola. We read (or are forced to study) their works, admire their heroes, and learn. In the case of Vasari, however, there's much more. We are transported into the minds and hands of great artists, and sit transfixed as we watch the unfolding of both history and genius — an apt comparison, I think, for the newly published work of Abraham 'Bram' Pais, albeit one he himself would not make.

Pais is a scientist of world class, and he has obviously been moving in the right circles for quite a few years now. He, too, has Vasari's capacity for making his biographees jump out of the page, both in his more extended works (like his famous ones on Albert Einstein and Niels Bohr) and within the 'microbiography' genre which he himself, I submit, invented.

Pais is the man who can say that he knows “so well” the article that Bohr wrote for Einstein’s seventieth birthday (1949), because “I helped him prepare it”; or the personal recollection of Paul Dirac (Nobel prizewinner at

31), and his invention of the 'bra' (the left half of an expression contained within a bracket); or his referring to transformation theory as "my darling".

And when did Pais first meet Wolfgang Pauli? At a 1946 party in Bohr's home, of course. The whole chapter on Pauli is so engrossing as to become unstoppable, even troubling reading: how could Pauli in 1921 (aged only 21) publish a review of the relativity theory that was so good as to merit Einstein's praise? Easy: he started early, during the boring hours at the *Gymnasium*, when he surreptitiously read Einstein's papers under his desk. We follow the development of Pauli's thoughts on physics, notably on quantum physics and the relations between spin and quantum statistics. A lifelong interest of his, which led to that pillar of physics known as the spin-statistics theorem, was, sadly, to be formulated only in 1958, the year of Pauli's death.

Indirectly, yet profoundly, we sometimes get the perspective of Pais himself, going well beyond personal recollections. During the description of Eugene Wigner's life, for example, the author admits to a difficulty in "grasping the personalities" of Hungarian-born physicists such as Leo Szilard, Edward Teller, John von Neumann (officially Margitai Neumann Janos Lajos) and, of course, Wigner himself. There is a quaint, lightly exotic something that sets the Hungarians apart from the more classic Mittel-European variety of the twentieth-century physicists who changed the world for ever.

There is a wider lesson, in history and sociology, to be learnt here, especially for younger readers. Of the 17 biographies in Pais' book, 13 are of European-born (and educated) scientists. Some, such as Isidor Rabi, are from countries like Galicia, which shifted from being a province of the Austro-Hungarian Empire to becoming part of post-Second World War Poland; some, like Pauli, simply say "I am European".

Pauli had his reasons. He was born and educated in Vienna, studied in Munich, was appointed ETH-Zürich full professor at 28, became, by default, German after the Anschluss (1933), but, in 1938, was denied Swiss citizenship because of his insufficient knowledge of "*Schwitzerdeutsch*" (while, of course, his *Hochdeutsch* was impeccable). Yet, after the war and the Nobel prize, he turned down lucrative appointments in the United States to come back to Zürich and, finally, to become a Swiss citizen. Of the other Pais personae, two are Chinese-born (Tsung Dao Lee and Chen Ning Yang), and only two, Mitchell Feigenbaum and Robert Serber, were born in the United States.

What better way to depict, besides the lives and science of the famed, the twentieth-century wave of migration of European culture towards the United States. This was a phenomenon of staggering proportions and extent. A different, but also very interesting,

account of the same global phenomenon is in a lesser-known book by Laura Fermi, *Illustrious Immigrants* (University of Chicago Press, 1968), which includes physicists, of course, but ranges also to other sciences, and to music and the arts.

In covering a wide range of topics, from relativity to chaos theory, Pais' life-and-science stories provide, above all, the staple on which young (and not so young) scientists should feed their ambitions. Alternatively, we can just enjoy and be enriched, both culturally and emotionally. ■

Giovanni F. Bignami is at the Agenzia Spaziale Italiana, Viale Liegi 26, 00198 Rome, Italy.

Wacky star of the Renaissance

Cardano's Cosmos: The Worlds and Works of a Renaissance Astrologer

by Anthony Grafton

Harvard University Press: 2000. 284 pp.

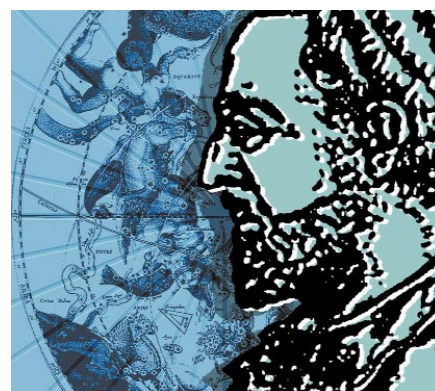
\$35, £21.95

Owen Gingerich

Girolamo Cardano was seen by his 1950s biographer, Oystein Ore, as "the gambling scholar", the man who introduced ideas of probability to Western civilization. Cardano's *Ars Magna*, the book containing the general solution of the cubic equation and generally seen as the greatest mathematical treatise of the sixteenth century, has been characterized as the first decisive advance over the attainments of the giants of classical Greek mathematics. The *Dictionary of Scientific Biography* lists medicine as the first of Cardano's qualifications and declares that among the physicians of Europe, he ranked second only to Vesalius. Yet another Cardano emerges in this new book: the leading astrologer of his day, rivalled only by Luca Gaurico and later by the Elizabethan magus John Dee.

It was no doubt inevitable that historian Anthony Grafton would bring his formidable intellect and lucid style to bear on Cardano. The Princeton professor has already completed a magisterial, two-volume study of the great Renaissance classical scholar Joseph Scaliger, whose father's greatest claim to fame was, in Grafton's words, "the most savage book review in the bitter annals of literary invective", a 900-page rebuttal of Cardano's *On Subtlety*.

Born in Italy in 1501, Cardano set out on the path to becoming Europe's most famous astrologer with the publication of an accessible astrological handbook in 1538. Five years later he achieved international notoriety when it was reissued as *Libelli Duo* in an enlarged version with 66 genitures, or birth horoscopes, of notable people. The book was



DAVID NEWTON

widely circulated, perhaps as much for its gossip and character assessments of the rich and famous as for its astral clues. That was in 1543, the year that the same publisher — Johannes Petreius of Nuremberg — brought out Copernicus's *De Revolutionibus*.

Two years later Petreius printed Cardano's most enduring book, his *Ars Magna*, an event off Grafton's map, although he does include a great deal of other fascinating information about the Nuremberg publisher. In 1547 Cardano produced a further expanded book of genitures, now *Libelli Quinque*.

Cardano's astrological publications had a variety of consequences. Now a controversial figure, he also became a sought-after adviser who could both prescribe medicine and deliver prognostications. He lived in a world of contradictions, where the messages from medicines and from the stars were fraught with ambiguity. It is this complex world of Renaissance growing pains that Grafton carefully dissects, with all its mixed signals and complexities. And he does it with literary grace and marvellous turns of phrase: "Cardano promised those who walked the paper corridors of his collection the secrets of both natural and political power". His acquaintance Reticus, Copernicus's only disciple, was none too pleased at being portrayed — in Grafton's vivid description — "as Dr Watson to Cardano's astrological Mr Holmes". In an "unbuttoned" autobiography Cardano portrayed himself as a "wacky professor", Grafton writes, considering this "a partial mirror of his fractured self".

Elsewhere, Grafton has written engagingly on the history of the footnote; here he reveals himself to be an erudite master of endnotes. A rich array of esoteric sources never hinders the narrative, but provides the evidence at the end of the volume. And when the desired materials aren't available from Cardano, because his letters had been burned, Grafton switches effortlessly to a well-documented case of how to give economic advice to clients by the infamous seer Nostradamus. Only very rarely can one catch Grafton out. During a meeting with John Dee in 1552, when Cardano had come

to Britain to treat the Archbishop of St Andrews for asthma, Grafton misidentifies "Joannes Franciscus" as Sir John Cheke, whose horoscope Cardano had cast, rather than as Jofrancus Offusius, an astrologer temporarily living off Dee's patronage.

For the overwhelming majority of *Nature's* readers, astrology has been relegated to the rubbish bin of history. Of what use is all this exotic scholarship? Let Grafton speak: "A knowing and thinking subject, Cardano modified, manipulated, or invented contradictory languages as he needed them. He was not always sure why he used them as he did, or on what assumptions they rested, or even whether the predictions he spoke came to him from his own intellect, the stars, or daemonic inspiration. But he knew and said that he stood at the center of a cosmos that his own mind could only mirror in a fractured, incoherent way. He deserves to be heard." ■

Owen Gingerich is at the Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA.

Not just a pretty Polly

The Alex Studies: Cognitive and Communicative Abilities of Grey Parrots

by Irene Maxine Pepperberg
Harvard University Press: 2000. 434 pp.
\$39.95, £24.95

John C. Marshall

From King Solomon to Walt Disney by way of Saint Francis of Assisi, the desire to converse with other species has run deep in the human psyche. The attraction furthermore appears to split we humans into two camps: those who wish to learn the 'language' of animals and those who are determined to teach them ours. On the whole, the former group has scored more scientific victories than the latter. The decipherment of bee dancing, for example, was a far greater achievement than any number of overexcited translations of chimpanzees waving their arms about in a bad imitation of American sign language. Worse, attempts to teach spoken English to apes was even less successful. The entire line of research in which we attempted to impose our communication system on another species was, it seemed, dead. The undertaking had run down the curtain and joined the choir invisible.

Not so. In *The Alex Studies*, Irene Pepperberg summarizes the first two decades of a project to communicate with African grey parrots (*Psittacus erithacus*). Pride of place in this enterprise is given to the Alex of the title. In some ways, this research programme has the look of a rational pursuit: parrots,

Pepperberg points out, have "a complex social system and a long life" and, more crucially, they do seem to 'mimic' human speech moderately well. On the debit side, many studies of mimetic birds have conspicuously failed to establish two-way speech communication with humans in properly controlled laboratory settings.

Pepperberg argues convincingly that the kinds of classical and operant conditioning techniques deployed in those prior investigations of mynahs, magpies and parrots were not well matched with the highly interactive, social manner in which mimetic birds acquire calls and songs in their natural environment. Accordingly, she modified an earlier technique of Dietmar Todt's, in which "humans assume roles played by psittacine peers in the wild". In Todt's original procedure, "one human is exclusively the principal trainer of each parrot, asking questions and providing increased visual and vocal attention for appropriate responses. Another human is exclusively the *model* for the parrot's behavior and simultaneously the parrot's *rival* for the attention of the principal trainer." In these circumstances, Todt's parrots learned relatively quickly (within a day) to mimic phrases produced by the 'rival'. But this behaviour is more akin to 'antiphonal duetting' than to functional communication, and would not surprise any parrot owner.

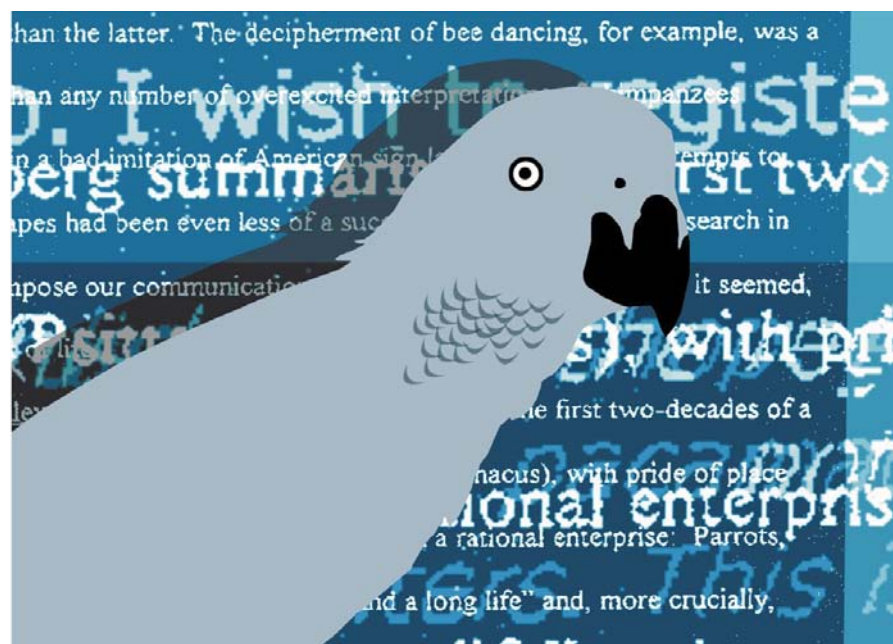
In Pepperberg's extension, one human 'trains' another to name objects, and to describe their colour, or shape, or number. The bird watches and is then brought into the game as the second 'pupil'. The two humans sometimes reverse their roles of trainer and trainee, and the bird is occasionally included in three-way interactions. Thus the birds "do not simply hear stepwise vocal duets, but rather observe a communicative process that involves reciprocity". Furthermore, the bird

can "effect environment change" (that is, obtain the object) by vocalizing "Wanna ..." plus some word in its vocabulary. The evidence presented suggests that Alex could acquire vocal labels for objects, actions and attributes to a level similar to that reached in earlier experiments with great apes taught to deploy visual signs or tokens.

Some generalization to novel situations is found, along with a very limited combinatorial ability. There were also indications that Alex can respond appropriately to sameness, difference and absence. Comprehension of relative size was fair, and that of object permanence was good, comparable to that demonstrated in young children by the Swiss developmentalist Jean Piaget.

For my money, the most interesting chapters were those concerned with just how similar (or dissimilar) to human speech is the parrot's imitation, and how the bird actually produces these sounds with a vocal apparatus that is very different from our own. In a sense, these chapters tell us as much about humans as about parrots — however distorted the signal, we are predisposed to hear speech. The final chapter includes fascinating information about the social behaviour of mimetic birds in the wild, and makes one wish for an entire book on this topic.

With respect to captive parrots, readers will have to make up their own minds about whether it was all worth the effort (on either Alex's part or that of Pepperberg). I suspect that those who were unimpressed by the proto-linguistic skills of chimpanzees and dolphins will remain undazzled by those of parrots. But people who thought these earlier attempts at interspecies communication were a major breakthrough in evolutionary science will no doubt be delighted with Alex's achievements and untroubled by how far back one must go to find a common ancestor



DAVID NEWTON

of parrots and humans. On the credit side, Pepperberg gives an excellent justification for her training procedures, the descriptions of the results are detailed and lucid, and, best of all, the interpretations are calm and considered. Pepperberg is well aware of Occam's razor and is careful not to indulge in ridiculous overinterpretations. One might nonetheless wonder what exactly was the point of this labour of love. To misquote Ludwig Wittgenstein, what would a parrot tell us if it could talk? Not a lot, seems to be the answer. Is a research programme to teach human speech to parrots likely to lead to deeper insights than one devoted to teaching us the calls and songs of the parrot?

Pepperberg's justification is that her data might help to improve the lives of captive parrots, "prevent habitat destruction and capture of birds in the wild, or enable researchers to develop better animal models for various human dysfunctions". I hope she's right. ■

John C. Marshall is in the Neuropsychology Unit, University Department of Clinical Neurology, Radcliffe Infirmary, Woodstock Road, Oxford OX2 6HE, UK.

Communications from the dead

Dear Mr Darwin: Letters on the Evolution of Life and Human Nature

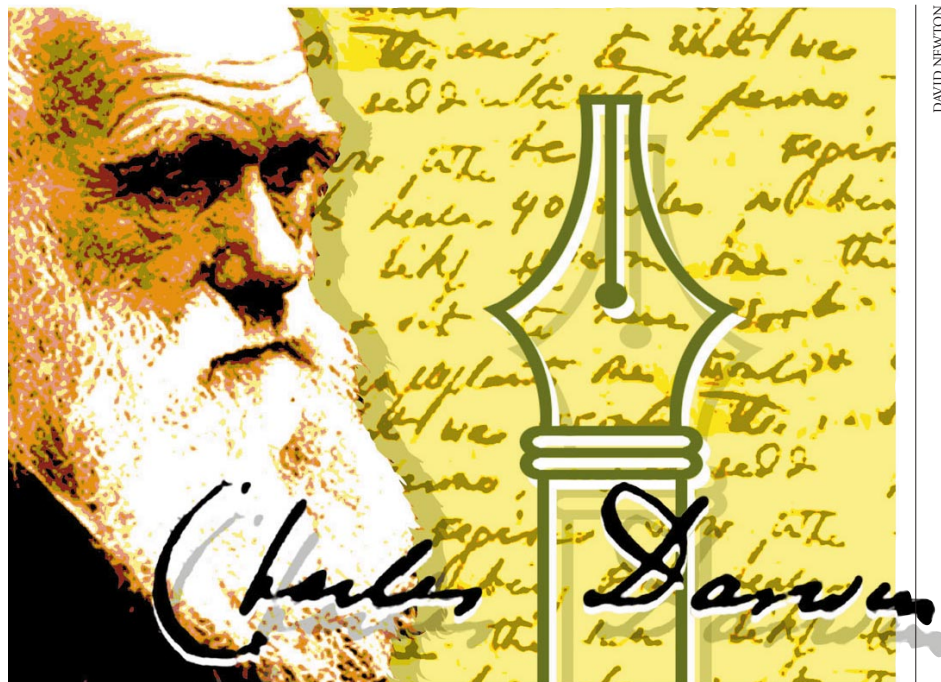
by Gabriel Dover

Weidenfeld & Nicolson: 2000. 268 pp. £20

A. J. Berry

In 1876 Charles Darwin contributed £10 — a substantial amount at that time — to the costs of the criminal prosecution of Henry Slade, a renowned spiritualist medium. Slade, his accusers charged, was a fraud, and his séances were merely elaborate exercises in legerdemain. Remarkably, the case pitted the two discoverers of natural selection against each other: Alfred Russel Wallace, author of an approving book on spiritualism, was the defence's star witness. Despite Wallace's characterization of the defendant as an "earnest inquirer after truth in the department of Natural Science", Slade was convicted. Darwin was delighted; he had no time for the "clever rogues" who preyed upon grieving relatives anxious to contact a loved one.

Darwin, who died in 1882, may now have cause to reconsider his attitude towards posthumous communication as he himself has recently taken to holding forth from beneath the flagstones of Westminster Abbey. The medium in this case is geneticist Gabriel Dover, whose book, *Dear Mr Darwin*, comprises a series of letters between Dover and Darwin. Dover brings Darwin up to date on



evolutionary biology since 1882, and Darwin, for his part, supplies appreciative yet inquisitive responses.

Things start rather formally — it's "My Dear Dover . . . Ever your most truly, Charles Darwin" to begin with — but become increasingly chummy as the correspondence develops — it's "Dear Gabby . . . Your most sincere friend, Chas. Darwin" by the end. The gimmick is almost painfully cute, but Dover handles it deftly: he is not unduly deferential, and his Darwin not overly impressed by what Dover has to say. The result is a quirky but readable account of the Dover perspective on modern evolutionary biology.

Darwin's education, however, is in idiosyncratic hands. At the outset, Darwin must predictably swallow doses of Mendel and Hardy-Weinberg, but the textbooks are then quickly forsaken when, on the second page of Dover's second letter, we run into his pet theory, 'molecular drive'. This, Darwin learns, is, along with natural selection and genetic drift, one of "the three forces of evolution". Much of the book is dedicated to explicating molecular drive and to justifying its exalted place in Dover's pantheon of evolutionary forces.

Dover introduced the term in the early 1980s after DNA-sequencing studies of multi-gene families — groups of related genes that often sit side by side along chromosomes — had revealed a striking and unexpected evolutionary pattern now known as 'concerted evolution'. Within a species, all members of a gene family may be identical, or at least very similar, whereas between even closely related species we see plenty of sequence divergence between homologous gene families. The homogenization of gene-family members within species is caused by a number of simple and well-understood

genetic processes, primarily unequal crossing over and gene conversion. Molecular drive is, in Dover's words, an "umbrella term" covering these and other "non-Mendelian mechanisms of inheritance".

Does molecular drive really rank beside selection and drift as one of the primary determinants of evolutionary change? Hardly. Darwin distinguished between two fundamental aspects of the evolutionary process: the genesis of variation, and the subsequent fate of that variation. In creating new configurations of existing genetic variation, molecular drive definitely contributes to step one. But does it contribute to step two? In principle, a variant can indeed spread through a gene family by molecular drive, especially when there are asymmetries in the drive process. For example, gene conversion is sometimes 'biased' such that an *a* allele is more likely to be converted to an *A* than an *A* to an *a*; such a situation may result in a molecularly driven increase of the *A* allele.

But crucially, the ultimate fate of any variant, whether subject to molecular drive or not, is determined by its impact on fitness: natural selection will intervene if it either enhances or diminishes its bearer's chance of reproduction. If the variant has no such impact — it is selectively neutral — then genetic drift is usually the major player, although molecular drive may sometimes also play a role. Molecular drive's contribution to the second phase of the evolutionary process is thus subordinate to the 'traditional' forces determining the fate of genetic variation in natural populations. Molecular drive is an interesting evolutionary phenomenon, but it is false advertising to bill it as a third major force of evolution.

Dover's Darwin, whose critical facilities

may have been dulled by a century or so underground, is more readily convinced of molecular drive's significance than I am. Having scripted Darwin's endorsement of his theory, Dover then settles down to enjoy his new role as Darwin's speech-writer. Responding to a lengthy Dover diatribe against Richard Dawkins, whose "selfish genery is genetically misconceived, operationally incoherent and seductively dangerous", Darwin reports that he will conscientiously hunt up Dawkins's books in a library: "I hope they are not filed under 'Science!'".

Dear Mr Darwin, however, is not confined to molecular drive and having Darwin say nasty things about Richard Dawkins. Dover writes at length on recent advances in developmental genetics, and adds his voice to those objecting to evolutionary psychology's insistence on attributing every quirk of human behaviour to the action of natural selection. Given that evolutionary psychology is an implicitly genetic theory (a trait must have a genetic basis to be subject to natural selection), it is interesting to note that many of its most persistent critics are geneticists.

Dear Mr Darwin is an engaging tour of Dover's passions, even if some are announced with more fanfare than they merit. Let us hope, however, that Dover's communications with Darwin do not create a literary fad based on the harassment of dead scientists.

The thought of Linnaeus being badgered by manic modern cladists is alarming. On receiving one of Wallace's spiritualist publications, T. H. Huxley replied, "I never cared for gossip in my life, and disembodied gossip, such as these worthy ghosts supply their friends with, is not more interesting to me than any other." ■

A. J. Berry is at the Museum of Comparative Zoology Labs, 26 Oxford Street, Cambridge, Massachusetts 02138, USA.



On being human

The Cultural Origins of Human Cognition

by Michael Tomasello

Harvard University Press: 1999. 248 pp.
\$29.95, £18.50

Andrew Whiten

Human language and thought elevate us mentally to a grade far removed from anything known in other animals. Yet it has happened in just a twinkling of evolutionary time. Less than six million years separate us from the non-human, non-verbal ancestor

we share with chimpanzees. So, suggests Michael Tomasello, we are faced with a puzzle: how could human minds vault this high so quickly? The question becomes more acute if one acknowledges little sign of any accomplishment beyond basic ape mentality until two million years ago or even less. Tomasello's solution — given how far he wants to push the idea — is a radical one. Depending on the reader, I suspect it will elicit excitement, irritation or incredulity. These different reactions may be more or less appropriate according to the evolutionary timescale Tomasello truly aspires to address.

The key proposition is that there was just one critical step in biological evolution which transformed our ancestors' capacity to sustain culture. A new 'ratchet' effect arose, in which cultural advances were built upon progressively in a way not seen in the social traditions of other animals. Human cognition would thenceforth become increasingly complex and differentiated, eventually achieving modern levels of sophistication without need of further biological change.

To see how radical a proposition this is, consider the case of language. Tomasello is arguing that the structures of our highly elaborated language capacities today have nothing to do with the evolution of a dedicated and, in some views, highly structured language instinct (he gives short shrift to the idea of innate mental modules, of any kind). Instead, he proposes that syntax and all the other complex aspects of human language have simply been built up over the generations, by cumulative, ratcheted, cultural evolution. The biological substrate did not need to change. What makes such a scenario plausible, according to Tomasello, is evidence that, quite early on, the human child demonstrates a capacity to translate between the perspectives of self and other that goes beyond anything seen in apes. Perceiving others as intentional agents, in particular, permits the child to become the kind of 'imitation machine' needed to participate in the powerful ratchet effect.

Tomasello's ambitious thesis requires accounts of changes on three very different timescales; evolutionary, historical and ontogenetic. To this task he brings almost unrivalled authority, based on an influential suite of both comparative and developmental studies; he cites more than 40 observational and experimental studies conducted by his group on monkeys, apes and children. These studies, mostly conducted in the 1990s, cover an impressive array of socio-cognitive capacities, including imitation, joint attention, theory of mind and language acquisition.

This substantial empirical base is coupled with a sophisticated grasp of the theoretical issues at stake, particularly when it comes to Tomasello's prime area of expertise, the development of language. Written with refreshing simplicity and directness, the

product is a slim volume that nevertheless packs in a richly articulated and challenging model of mind, backed by a wealth of pithily summarized comparative and developmental studies.

At least two-thirds of the book is devoted to tracing the origins and development of components of cultural learning in children, with a particular emphasis on language. This is a masterly survey, covering pre-linguistic scaffolding for language, the acquisition of symbol and syntax use, discourse and the implications of internalization for other aspects of cognition.

Certain features of Tomasello's thesis are less compelling. He considers the possibility that the vital change may have happened two or even six million years ago. But his argument appears to neglect enormous changes in the brain, which has tripled in size since six million years ago and roughly doubled in the past two million. It seems more likely that whatever elaboration of social and cultural practices occurred in this period, it was underwritten by equally massive and rapidly driven neural changes.

A predominant role for cultural change becomes more likely in the context of the past quarter of a million years of *Homo sapiens*' existence. If the greater part of existing language structure arose over this period, the idea that this happened through cultural learning and ratcheting processes still constitutes a major challenge to those who argue for innate language systems. Tomasello notes that the main diversification of the Romance languages occurred in a few hundred years; so why could not cultural processes of syntacticization turn an embryonic language into a vastly more complex one over hundreds of thousands of years?

A further doubt is whether Tomasello has correctly identified the critical cognitive step that elevated our ancestors' social sophistication over existing anthropoid psychology. His conclusion is largely founded on experimental findings in captive apes far removed from the rich inputs of their natural environments. Field researchers tend to perceive more advanced cultural processes at work, although these perceptions are difficult to substantiate without experimental controls. Accordingly, we are at something of an impasse on this question. Tomasello's thesis probably depends less than he implies on the exact difference between chimpanzee and human cultural propensity, especially if the thesis gets its main application in the recent rise of *Homo sapiens*.

Nevertheless, students of primate behaviour are one of several groups who should read this important book. It spells out forcefully what appears to make human development so distinctive, and does so from the perspective of an expert in language acquisition who has also devoted much time to comparative work with apes. It is strong medicine for

anybody in danger of romanticizing the similarity of ape to child. Developmental psychologists will find here a well-articulated account of the ontogeny of cultural learning, which challenges alternative accounts from the vantage point of extensive research. The book should also be a thought-provoking read for cognitive psychologists, many of whom seem to disregard social and cultural processes. Of course, we have 'cultural psychology', but that is about cultural differences. Tomasello is instead talking about universals, an aspiration of most cognitive psychology. If he is even half-right, the field suffers a serious omission indeed. ■

Andrew Whiten is at the School of Psychology, University of St Andrews, St Andrews, Fife KY16 9JU, UK.

The sequence of words

Genes, Peoples, and Languages

by Luigi Luca Cavalli-Sforza

Farrar, Straus & Giroux: 2000. 216 pp. \$24

Patrick V. Kirch

In the early twentieth century, a uniquely American version of anthropology emerged under the direction of prominent scholars such as Franz Boas, Alfred Kroeber and Edward Sapir. What set this school apart from its European counterparts was its efforts to construct a scientific approach to deep-time human history, and the catholic inclusion of evidence from ethnography, archaeology, linguistics and human biology (then called physical anthropology). Today, tensions within anthropology threaten to explode this model of a 'holistic' approach to the study of human cultures. Luigi Cavalli-Sforza, a geneticist whose research has depended on close collaboration with archaeologists and historical linguists, gives us ample reason to celebrate anew this holistic vision of anthropology.

For several decades, Cavalli-Sforza and his colleagues have pursued a variety of research projects involving not only the mapping of human genetic variation, but also the correlation of geographic patterns of variation with patterns of language distribution, and with archaeological evidence for ancient population movements. The underlying theme integrating these disparate studies traces the biological and cultural evolution of modern *Homo sapiens* throughout the past 100,000 years.

Over the course of Cavalli-Sforza's career, the methods for studying genetic variation have themselves been radically transformed, from the relatively crude mapping of ABO blood groups, to the sequencing of DNA, and even the recovery and sequencing of ancient



DAVID NEWTON

DNA itself. *Genes, Peoples, and Languages* tells the story of this research odyssey in a style that is highly accessible to intelligent readers from any background, without presuming prior grounding in genetics, statistics or, for that matter, linguistics and archaeology.

In his effort to make his research methods transparent to those without a background in human genetics, however, Cavalli-Sforza ends up devoting much of the first half of his book to a kind of primer in genetics, evolutionary trees and multivariate statistics. This may deter some readers, for the fascinating problems of deep-time human history do not begin to emerge until the fourth chapter. But for those who persist — or simply skip ahead — the intellectual rewards are ample.

In the chapter "Technological revolutions and gene geography", Cavalli-Sforza tackles one of the great problems of historical anthropology: the expansion of modern humans out of Africa, and subsequent diasporas that correlated with major periods of demographic growth. Here he draws on his work with archaeologist Albert Ammerman in demonstrating a major population-expansion-driven (as opposed to strictly cultural) diffusion from the Near East into Europe, associated with the spread of agricultural peoples. But while the genetic landscape of modern Europe still bears the signature of this major diaspora, other 'streams of gene diffusion' are also evident in statistical patterns of gene distribution, including an expansion from the region north of the Black Sea. This appears to have been associated with the expansion of pastoral nomads who originally domesticated the horse.

Although he dwells at greatest length on Europe, Cavalli-Sforza also discusses evidence for several other major diasporas, including the expansion of Bantu language-speaking groups in sub-Saharan Africa, and the remarkable dispersal of the Austronesian-

language speakers. This latter dispersal would ultimately result in the human colonization of the most remote Pacific islands, including Easter Island and Hawaii.

There are, of course, always pitfalls in multidisciplinary research waiting to entrap unwary scientists stepping beyond the familiar terrain of their own discipline. Writing on language, Cavalli-Sforza is liable to raise some hackles, for not all of his claims fall within what is currently accepted by mainstream historical linguists or prehistorians. This derives, in part, from Cavalli-Sforza's desire to have a 'big picture' classification, or family tree, of world languages, which can then be correlated with his genetic tree of human populations. Unfortunately, the classic methods of historical linguists have serious limitations once one attempts to move back in time to a level deeper than that of major language families. Cavalli-Sforza thus relies heavily on the work of linguists Joseph Greenberg and Merritt Ruhlen, especially the latter's highly controversial 'super-family' classification.

For example, in the key diagram comparing genetic and linguistic trees, Cavalli-Sforza identifies a 'Melanesian' population, and groups it genetically most closely with 'Polynesians' and 'Micronesians', and more distantly with a 'New Guinean' population. Linguistically, the Melanesians and New Guineans are depicted as linked together in an 'Indo-Pacific' family, while the Polynesians and Micronesians are identified as part of the large 'Austronesian' language family.

However, the very concept of a discrete 'Melanesian' population was rejected many years ago, and the notion of an 'Indo-Pacific'-language family was similarly abandoned. Rather, the people who occupy the islands of Melanesia (excluding New Guinea) display tremendous genetic variation. This is the result of a specific and complex history of

human migrations, including an early phase of hunter-gatherer expansion around 40,000 years ago, and a later phase of population intrusion by Austronesian speakers about 3,500 years ago — reflected archaeologically in the Lapita cultural complex. This latter intrusion resulted in virtually all island Melanesian populations speaking Oceanic languages (a branch of Austronesian).

Other regional specialists will doubtless find similar problems with Cavalli-Sforza's efforts to correlate genes and languages at such a broad level. But this does not mean that the effort to construct such worldwide correlations is futile, for there is clearly much historical truth in Cavalli-Sforza's models. As a first approximation, the history of human diasporas painted in this book shows a vast improvement in our knowledge, which has been made possible by rapid advances in method. And not just in biological science, for the historical linguists themselves are now tackling the thorny issues of deep-time language reconstruction more intensively than ever before. My point is simply that Cavalli-Sforza's models will need more work, and much refinement, before they will satisfy all the critics.

In his final chapter Cavalli-Sforza ranges widely over some of the most profound questions of human evolution, including the role of culture as a means of biological evolution, the modes by which culture is transmitted, and the emergence of complex hierarchical social structures. The ideas raised here reinforce how essential it is for social scientists and biologists not to be isolated — as they so often are in the intellectual structures of modern academia — and to realize how much they need each others' insights.

Edward Sapir eloquently laid out a multidisciplinary methodology for historical reconstruction in his now little-read 1916 monograph *Time Perspective in Aboriginal American Culture: A Study in Method* (Canada Department of Mines). At the time, the most sophisticated discipline contributing to the tracing of deep-time human history was linguistics. Sapir himself lamented that he had probably "undervalued the data of physical anthropology". Little could he have anticipated the advances that would come with the revolution in molecular biology and genetics.

Cavalli-Sforza and his colleagues have now given historical anthropology sophisticated tools to look at human variation, and to read that variation as a 'text' of human history. Cavalli-Sforza writes that "to some, history is not a science", but he has helped to show that history is susceptible both to the methods of science and to the power of a multidisciplinary approach. ■

Patrick V. Kirch is at the P. A. Hearst Museum of Anthropology, University of California, 103 Kroeber Hall, Berkeley, California 94720-3712, USA.

Evolution in a broader perspective

Lucy's Legacy: Sex and Intelligence in Human Evolution

by Alison Jolly

Harvard University Press; 1999. 518 pp.

\$29.95, £18.50

Christophe Boesch

Gigolo in French means a male prostitute, while in German it may mean a Don Juan, and some Americans see in it a man used by women in one way or another. Alison Jolly probably had this last sense in mind when she provocatively proposed to rename a group of primates with many females and a single male a 'gigolo' group. The alternative term 'harem' suggests that it is the male who is most important; if he's a gigolo, he's there because the females tolerate him.

Alison Jolly's book is a refreshing and stimulating account of evolution, and especially of the evolution of sex and intelligence in humans and primates. One of its strengths is her presentation of the female view. We men have always had a tendency to consider life mainly from our own perspective. Thus, terms such as 'female choice' or 'female control' rapidly faded out of favour despite the fact that Darwin used them. But it is important for men to consider the other view of evolution, sex and intelligence — especially since there are good theoretical reasons to expect that intelligence can out-compete strength, and that internal fertilization gives every female mammal the final choice of which sperm is going to fertilise her egg. It would be surprising if females did not use this huge advantage.

Jolly proposes that the traditional feminine viewpoint on evolution is one of cooperative organization and not of competition. This is a counterpart to the fundamental dilemma for anyone trained in Darwinian evolutionary theory, with its emphasis on rampant individualism: how did society develop in the face of this? For her, "selfish

genes, interacting with their environment, led to love between kin, trust between friends, the intricacies of the mind, and the emergent organisations of society".

Jolly tells the story from her own perspective as a woman, teacher and lemur-watcher, taking examples from women's lives such as female orgasm, menstruation, childbearing and menopause. Some of these stories are quite illuminating. To explain to sceptics that not all human behaviour is either learned or conscious, she takes the example of labour. "Your body, willy-nilly, starts to sneeze, goes into labour, opts to move from first- to second-stage contractions. A nurse told me once, 'Don't you dare push!' 'I'm not pushing!' I squawked, 'It is!' ... the body's wisdom during labour does not rest on your conscious decisions."

Obviously only men could propose in face of such evidence that human behaviour is a *tabula rasa*! The only regret I have is that her case for showing that evolution is as much about cooperation as about competition is not convincingly presented. Obviously cells in the same organism must cooperate, otherwise they would commit suicide. But individuals do compete regularly and Jolly does not present adequate evidence for cooperation.

A good case could have been presented about sex. Do males and females cooperate or compete? While commenting that "politically correct biologists try hard to describe sex as cooperation between equal partners", Jolly presents evidence of both competition and cooperation. In the course of evolution, the two sexes have differentiated in such a way that they depend on one another, so that sex does not work without at least some cooperation. But for readers who fear that biologists might apply what we observe in animals to human beings, she rightly says of observed instances of violence and killing in animals that "what is natural is not necessarily right", leading the way to more philosophical questions. As debate rages about the existence of 'moral' actions in animals, she tells us that, for some people, "it seemed immoral that morality should increase genetic fitness".

In other words, this book is a quest for a better understanding of human evolution and a challenging contribution to the debate about human uniqueness, if this does indeed exist. I strongly recommend it to those who are new to the field because of the breadth of the topics and the clarity with which they are presented. To those working in the field I strongly recommend it for the new approach Alison Jolly takes to some of those questions. She suggests that beside concealment of ovulation in women, what is new in human sex is the general need for privacy. But I'll leave readers to discuss that for themselves. ■

Christophe Boesch is at the Max Planck Institute for Evolutionary Anthropology, Inselstrasse 22, 04 013 Leipzig, Germany.



How much is really changing in Japan?

Japanese Science: From the Inside

by Samuel Coleman
Routledge: 1999. 214 pp. \$90, £55

Robert Triendl

John Desmond Bernal, the noted crystallographer and vigorous social critic, had little good to say about Japanese science. His verdict, published in his book *The Social Function of Science* (Routledge, 1939), was devastating. For Bernal, Japanese science was hopelessly inaccurate, bureaucratic and misguided by political ideology. Since he was writing in 1939, though, his judgement is perhaps better understood as a reflection of antagonism towards an expansionist military power. In fact, by that time, Japanese scientists had made important contributions in various disciplines.

Japanese science has come even further since then. After almost three decades of darkness and severe material limitations following the Second World War, funding for research has soared. Today, Japanese laboratories are among the best-equipped in the world and there can be little doubt that Japan matters in virtually any field of science and technology. Yet it is not unusual to hear Japanese scientists describing the reality in terms that differ little from Bernal's harsh assessment of more than 60 years ago.

In *Japanese Science: From the Inside*, the American anthropologist Samuel Coleman provides a profound and insightful critique of scientific organizations in Japan. The book is based on extensive fieldwork in a number of bioscience-related laboratories and research institutes. And, most importantly, rather than rushing towards his own judgements, Coleman provides ample space for the views and voices of Japanese researchers themselves.

Complaints about bureaucratic rigidity abound in Japan, and the degree of frustration in some quarters, especially among young scientists, is alarmingly high. If one is to believe grass-roots polls undertaken in anonymity by academic societies, most scientists in the country believe that promotion has more to do with personal connection than with individual achievements, and that the evaluation procedures of the Ministry of Education's US\$2 billion university grants system is seriously flawed.

The vast majority of basic science is undertaken in national universities and it is there that most of the problems lie. Hiring and promotion practices remain largely obscure, access to resources remains under the control of a few senior professors at major



national universities, and administrative rules strangle university research.

It is telling that no Japanese university ever offered a professorship to Shuji Nakamura, who pioneered research on gallium nitride semiconductors while working for a local manufacturer of fluorescent materials. Japanese national universities remain 'closed shops' not only to foreigners and the growing number of female researchers, but also to anybody outside the academic networks of power. Japanese scientists who trained overseas often find it difficult to integrate into the Japanese academic community.

While Coleman's book is a devastating assessment of the scientific enterprise in Japan, he does not derive his conclusion from simple comparisons with elsewhere. Instead, he looks at disruptions to what Bruno Latour and Steve Woolgar in their study *Laboratory Life* have called the "cycle of credibility" (Coleman terms it the "cycle of credit"), the cumulative process by which research output is translated into resources for future scientific work. Coleman makes few assumptions about how efficient scientific organizations are meant to be, as long as there are no major distortions of the credit cycle of scientific work.

Coleman has done fieldwork in a number of settings — from an application-oriented research institute under the Ministry of Agriculture to a laboratory in the medical school of a national university. But he has spent most of his time at two neighbouring institutions in Osaka: the Osaka Bioscience Institute (OBI) and the Protein Engineering Research Institute (PERI), both considered major organizational innovations in Japanese public research.

PERI and OBI are unusual in that, from the start, only limited-term positions were offered. Although these institutions were intended as show-pieces of a changing Japan,

their limited-term policy created problems in the country's inflexible academic labour market, confronting scientists with the rigidities of the Japanese academic world in a sometimes shocking way. Most researchers, Coleman tells us, were there because they "had nowhere else to go" — they had unusual backgrounds or were pursuing research topics that didn't fit in with established academic career trajectories. At PERI and OBI they found a formidable research environment and unprecedented levels of support, but were locked into a somewhat unreal position that offered little except good conditions. Without much prospect of a position that would allow them to continue research, for some their assignment at PERI or OBI was little short of a moratorium.

Questions that affect the whole Japanese scientific community have taken on more articulate forms within these two institutions, but Coleman does not ignore conventional research settings. Novel arrangements such as those at PERI and OBI have made new career trajectories possible, although how much flexibility they have added to the Japanese public research system is difficult to evaluate. Considering their limited number, it is hard to reject arguments that they have simply weakened the demand for more profound reforms.

Japanese reformers are currently trying to increase job flexibility and prioritize funding. Coleman's book should be a warning to those who think that increased labour mobility in the lower ranks of the academic workforce will be sufficient to rework Japan's scientific enterprise. As long as other elements of the credit cycle — most notably evaluation and hiring practices — remain unchanged, the outcome could well be detrimental. ■

Robert Triendl is in the Tokyo office of the Asian Technology Information Program, 6-15-21 Roppongi, Minato-ku, Tokyo 106-0032, Japan.

Putting technology in its place

Why don't Europeans carry Mayan calendar calculators in their Filofaxes?

Don Ihde

Recently, multicultural and nonlinear accounts of scientific discovery and technological innovation have challenged the dominant linear, often eurocentric histories of science. It has become commonplace to acknowledge that the mathematical concept of zero originated in India, the magnetic compass was developed in China; and even moveable type came from Korea — all before they appeared in Europe.

But it has been harder to recognize non-European attainments as sometimes superior, and even harder to understand the barriers to intercultural transfers. Mesoamericans, for example, developed complex and accurate calendars, implying sophisticated measurements and observations. The Mayans even created a calendar calculator (in stone!) that was more accurate than contemporary European calendars — the Mayans did not need the Gregorian Reform which eliminated 11 days of 1572. The device that did the Long Count was not even recognized as a calendar until centuries after the Spanish Conquest. But the Mayan calendar has never been transferred. The reasons are easy enough to discern — technological devices are embedded in levels of contextual meanings, apart from which they are merely 'objects'.

The Mayan calendar calculator is embedded in a base-20 mathematics, a celestial cycle of 18,980 days, and a dating system that runs for four days. For the calendar to be read, a whole system of mathematics would have had to have been transferred. At a higher cultural level, calendars are needed to fix socially important events with regularity: Nile floods for the Egyptians; solstice for the Romans; Easter for Christians; and the regularities of the War Star (Venus) for the Mayans. Without an exchange of cultural significations, Mayan cultural meaning would also be difficult to transfer.

The magnetic compass has a long and complex multicultural history, but by the time it had been europeanized, it was but one instrument in a complex set of mathematical methods for mapping longitude and latitude, from which the navigator could tell where he or she was. Puluwateans, one of many South Pacific cultures, could determine their location by means of wave patterns, flights of birds, refractive effects from islands, and so on. The Puluwateans adapted the compass, largely as a fascinating object. But it could only be used to steer a straight course, and this placed the Puluwatean navigator in danger

of losing the skills by which one steered a straight course using regular wave patterns. The invention transferred, but the replacement of cultural context had negative results. The compass, moving from one culture to another, became something else.

A reverse transfer from the South Pacific back to Europe also occurred, but belatedly. Captain James Cook, during his famous expeditions of the eighteenth century, recognized the superior speeds, manoeuvrability and seaworthiness of the large Polynesian catamarans (some of which were as long as European vessels), yet multihull design was not adopted in the West until much later. This was probably due to the European need to carry heavy cargo (and cannons), something inimical to high-speed multihull design. And resistance also lay in the inability to transfer techniques from one society's builders to another. When Nathaniel Herreshof, the most notable nineteenth-century American sailing designer, did produce a few racing catamarans, they were quickly declared illegal by the racing clubs, as the innovation threatened the entire racing tradition. Today's multihull sailing

vessels have decimated all previous sail racing records, and high-speed catamaran ferries ply the world's waters.

By recognizing that a technology or an instrument is what it is in relation to its context of meanings, it becomes possible to take deeper account of the problem of recognition itself. The mostly recent and anachronistic interpretations of ancient astronomical instruments — for example, the range of claims interpreting Stonehenge, sighting tunnels in Mexico, and a whole range of related artefacts — as measurement devices, may be as much a reflection of our own obsession with measurement as not.

Nonlinear and multicultural interpretations of the history of science and technology, far from being — as eurocentrists sometimes claim — an opening to relativism, actually recognize the ingenuity of the entire species. The discovery of 800,000-year-old Acheulean-like stone technologies in Asia not only defeats the infamous archaeoracism of the Movius Line (which divided Asia from Europe and Africa; the lack of tools on the Asian side led to the belief that *Homo erectus* was more conceptually and technologically advanced in Africa and Europe), but pushes humankind's nonlinear, multicultural history further back in time. Contrarily, revisionist and nonlinear interpretations of science and technology are not only more sensitive to, but more respectful of human ingenuity wherever it is found. ■

Don Ihde is in the Department of Philosophy, State University of New York, Stony Brook, New York 11794-3750, USA.

James Cook recognized the superior speed and seaworthiness of the Polynesian catamaran.



Awkward date: Mayan calendars, such as this one, were more accurate, but used base-20 mathematics.

The Enduring test

How long must you wait before proving your humanity?

Paul Levinson

"How do I demonstrate my humanity to you?" Her blue eyes glittered. Her mouth grew taut. She looked vividly human, in every way.

"You cannot," I replied, carefully. Every word had to be well chosen. "Humanity is not something that can be demonstrated. It is rather something that is."

She sipped slowly from a cup of tea, regarded it, then placed it back on the table. "And what it is that makes an individual human? I presume you're not talking about a soul?"

I smiled. She was good. "I am not," I replied. "As far as I can tell, most people believe that possession of a soul is strictly in the eyes of the beholder."

"Very well," she said. "So we see things the same on the soul count. What else do you hold essential to the condition of humanity? Is it in the way that the DNA comes to be?"

"I doubt it," I replied. "What does it matter if DNA is spliced and sewn together, thread by thread, or emerges all at once from some unplanned shuffle of the chromosomes, if the result is the same? Anyway, DNA is a blueprint, a recipe. The truth of humanity is in the eating."

"Very nice," she said, and nodded her appreciation.

"Thank you," I replied.

"So what is it, then? Art?"

I shook my head no. "Painting by the numbers has been done since before there were personal computers."

"Music?"

"Not likely," I said. "Computers have been composing fine music for centuries. Mathematics and music have worlds in common."

"Then what? Sense of humour?"

"Try this," I replied. "A man is building a house. He has 25 bricks left. He completes the job with 24. What does he do with the extra brick?"

She shrugged. "I give up."

"He throws it in the street," I said.

She looked at me for a long second. "That's not funny. I don't get it."

"No? Well, let me try another one on you. A man is smoking a big, foul-smelling cigar on a bus. He's seated next to a woman who has a poodle on her lap — shedding hair, and barking its head off. The two glare at each other. Finally, the woman, unable to contain herself, rips the cigar out of the man's mouth, and tosses it out the window. The man retaliates: he scoops up the little dog, and throws

that out of the window. The two continue arguing, and get off at the next stop. As the bus pulls away, the poodle comes running up to the woman. It has something in its mouth. Know what it is?"

"The cigar?"

"No. The brick," I said.

She looked at me, then burst out laughing.

"So you see," I said. "It's not sense of humour either. You and I agree that the first — let us call it, sub-joke — was not funny. And the second, which fulfills the cycle, is at least worth a laugh."

She smiled again, then sighed. "So what is the essence of humanity? What is it that distinguishes you from me?"

"Have you heard of Alan Turing?" I asked.

"Of course," she replied. "The Turing test is hundreds of years old. If a non-human is indistinguishable from a human in the performance of its intelligence, then by what logic ought humans deny that the non-human has human intelligence?"

"Indeed," I said. "So, can you think of any such logic?"

She considered. "Well, we know that parrots do not really speak. Rather, they make sounds that mimic speech. And yet, according to the Turing test, one might listen to a parrot for a short time and conclude it was really speaking. But that can't be right. So there's something wrong with the test."

"Precisely!" I said. "Excellent. So, would you like to hear my addendum to the Turing test?"

Her eyes lit up. "I would."

"Well, I call it the Enduring test. It would go like this: if a non-human were indistinguishable from a human in the performance of its intelligence *forever*, then no one could deny that the non-human had human intelligence. This takes care of the parrot problem, you see, because even the cleverest parrot can only mimic for a brief time."

"Hmm..." she said. "Very good. I like that. But it seems a little extreme. Who could wait forever for the results of any test?"

"Yes, well, that's just the outer boundary, the ideal," I replied. "We could retreat to a lesser position and say, the longer a questionable entity was indistinguishable from a human in performance of intelligence, the less reason anyone would have for denying that such an entity had human intelligence. Would that work for you?"

She thought, then nodded her head slowly. "Yes, it might well. That's very good." She slowly pushed her chair back, and stood up. She extended her hand to me. "Thank you.



This conversation was *very* useful — one of the best yet. We'll be back to you on your application very shortly."

I looked at her through the window as she left. I very much wanted to be a school teacher. I could understand their reticence — their concern about leaving their children in even the temporary care of an artificial creation. Still, I knew I could do the job. I sighed. I guess I could last a hundred more years. That's what the specs said about my body and brain. But I didn't want to wait forever for their decision.

Paul Levinson's most recent books are the novel *The Silk Code* (Tor Books) and *Digital McLuhan: A Guide to the Information Millennium* (Routledge).

JACEY

Ringling in the new cosmology

Wayne Hu

Balloon experiments over Antarctica have produced a long-awaited temperature map of the microwave sky. The map reveals sound waves that can be used to probe the early Universe.

According to the theory of the Big Bang, the Universe started hot and dense and then expanded and cooled. In the hot, dense conditions of the early Universe, photons were tightly glued to matter. When the Universe was about 300,000 years old the temperature dropped below 3,000 K, allowing atomic hydrogen to form and releasing the photons. These photons, which travelled freely through the Universe as it expanded and cooled, make up the cosmic microwave background (CMB) we see today. Ten to twenty billion years after the Big Bang, the CMB is a cold sea of photons with an average temperature of 2.7 K (−270 °C). These photons are all around us, causing about 1% of the noise on our television sets.

When it was discovered in the 1960s, the CMB was found to be remarkably uniform across the sky. It was not until 1992 that the Cosmic Background Explorer (COBE) satellite¹ discovered temperature variations (or ripples) at the level of 1 part in 100,000. Temperature maps of the CMB form a snapshot image of the Universe when it was extremely young. So these ripples reflect tiny density fluctuations in the primordial soup of particles. These same density fluctuations are thought to grow by gravitational attraction into the familiar structures we see today (stars, galaxies and clusters of galaxies). This is the gravitational instability model of structure formation.

COBE told us what the large-scale fluctuations in the background look like, but cosmologists today are more interested in the small-scale fluctuations. Astronomers divide up the sky into angular degrees, so that 90° is the distance from the horizon to a point directly overhead. COBE measured temperature ripples from the 10° to 90° scale. This scale is so large that there has not been enough time for structures to evolve. At the degree scale, on the other hand, the process of structure formation imprints information in the ripples about conditions in the early Universe.

Since the COBE discovery, many ground and balloon-based experiments have shown that the ripples peak at the degree scale^{2,3}. On page 955 of this issue, de Bernadis *et al.*⁴ report the first high-resolution maps of the CMB over a significant part of the sky. The results are part of the BOOMERanG experiment, in which a microwave telescope was carried at high altitudes during a long balloon

flight over Antarctica. The authors take a power spectrum from their detailed maps of the CMB, much as you would if you wanted to analyse background noise. They find that the peak in the power spectrum of the CMB has exactly the right form to be the ringing or acoustic phenomena long awaited by cosmologists (Box 1). Such acoustic phenomena probe the conditions of the early Universe as a kind of cosmic ultrasound (Fig. 1, overleaf).

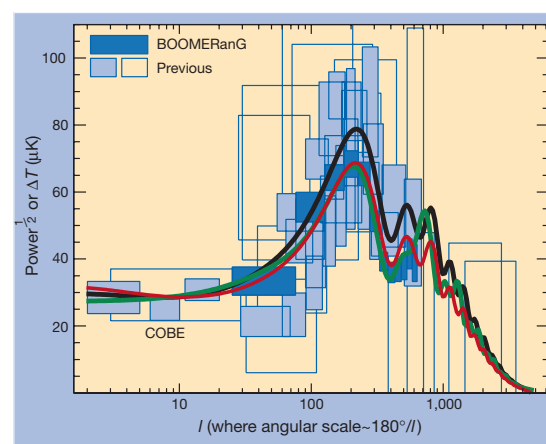
The properties of the observed peak have important implications for cosmology⁵. They depend mainly on the spectrum of initial fluctuations and fundamental cosmological parameters. The location and width of the peak strongly imply that the initial fluctuations that created the sound waves were in place on the largest scales at the earliest times. The only known mechanism for setting these perturbations in place is a

Box 1 The second peak in the spectrum

Before the BOOMERanG data⁴ discussed here, all the evidence pointed towards a model of the Universe that is flat and lightweight (low dark-matter density), with an initial spectrum of density fluctuations whose power is constant across all length scales⁶. This standard model (black curve) is strongly inconsistent with the observed lack of a prominent second peak in the power spectrum of the cosmic microwave background.

There are at least three possible explanations for the 'missing' peak. First, the initial density fluctuations could actually increase with length scale, thereby suppressing small-scale fluctuations. This is known as a 'tilted' model (red curve, 10% tilt). This solution would have important implications for the particle physics of inflation and observations of gravitational waves.

The second possibility is that the density of baryons (ordinary matter) is as much as 50% higher than the value implied by the abundance of light elements in the Universe and the theory of their synthesis in the first few minutes after the Big Bang (nucleosynthesis)⁶. Any extra baryons cannot be in the stars we see today. If this were the



solution, the question of where most of the baryons are today becomes even more puzzling⁷. At the very least, the BOOMERanG data place a lower limit on the baryon density that is comparable to the nucleosynthesis estimate.

A value for the dark-matter density higher than the standard one-third of the critical density also helps fit the power spectrum better (green curve; three times as much dark matter and 50% more baryons), at the expense of agreement with other cosmological data⁶. The standard value can also be made to work by lowering the predicted height of the peaks relative to the COBE measurements at the 10° scale by any one of several effects.

A testable consequence of either high-baryon-density solution is that the third peak should be higher in amplitude than the second.

The final and perhaps most speculative solution is if the formation of atomic hydrogen were to be delayed until the Universe was nearly 30% older, either through an unknown source of energy or through a change in our understanding of atomic physics at early times. This would increase the time available for the acoustic oscillations to dissipate and hence suppress the smaller peaks. A combination of some or all of these solutions may also provide the answer and perhaps avoid any extreme departures from the standard model.

W.H.

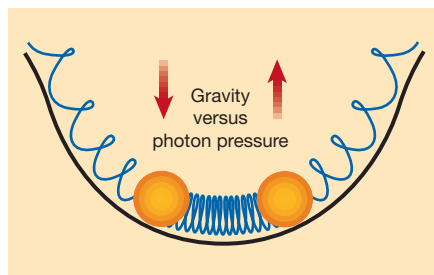


Figure 1 What astronomers see in the fine structure of the cosmic microwave background is actually sound waves. Acoustic phenomena have long been a general prediction of gravitational instability models of structure formation⁸. Gravitational force compresses the primordial plasma until resistance from photon pressure reverses the motion, leading to acoustic oscillations. Because compression raises the temperature, this results in hot and cold spots that are visible in the microwave sky today.

period of rapid (faster than light) expansion in the early Universe called inflation. In inflationary models, quantum fluctuations are carried from the microscopic to cosmic scale by the rapid expansion. If this interpretation is correct, then there should be a second peak that is smaller than the first, and a third peak that is comparable to or larger than the second.

In the inflationary context, the structure of the peaks is governed mainly by three fundamental parameters. These are the curvature of the Universe, the density of ordinary matter (or baryons), and the density of dark matter. The location of the first peak in the power spectrum provides the best measure of the curvature of the Universe, and hence the total amount of matter in the Universe. Einstein told us that matter curves space: the familiar force of gravity is no more than the curvature of space-time. To see this fact, consider the surface of the Earth. Two people travelling due north from the Equator on different lines of longitude will nonetheless meet at the North Pole. Ignorant of the curvature of the Earth, they might attribute this fact to a strange attractive force.

The same thing happens to CMB photons on their way to the observer if the Universe is spatially curved. The intervening matter and energy acts as a giant (de)magnifying glass that bends the photon trajectories. The BOOMERanG result supports a flat Universe, which means that the total mass and energy density of the Universe is equal to the so-called critical density. A perfectly flat Universe will remain at the critical density and keep on expanding forever, because there is not enough matter to make it recollapse in a 'big crunch'.

Perhaps the most intriguing aspect of the BOOMERanG results is the lack of a prominent second peak at half the angular scale of the first. If the second peak predicted

by inflation exists, it is of much smaller amplitude than the first. Indeed it must be significantly smaller than expected from the simplest models (Box 1). The key to resolving the mystery of the second peak will be measurements of higher precision and resolution, perhaps from the forthcoming full analysis of the BOOMERanG data — but, if not, certainly from the MAP satellite to be launched in November. Regardless of the outcome, these data show that we have clearly entered a new era of precision cosmology, in which we can begin to talk with certainty

about the origin of structure and the content of matter in the Universe. ■

Wayne Hu is at the Institute for Advanced Study, School of Natural Sciences, Olden Lane, Princeton, New Jersey 08540, USA.

e-mail: whu@sns.ias.edu

1. Smoot, G. *et al. Astrophys. J. Lett.* **396**, L1–L4 (1992).
2. Miller, A. D. *et al. Astrophys. J. Lett.* **524**, L1–L4 (1999).
3. Mauskopf, P. D. *et al. Astrophys. J. Lett.* (in the press).
4. de Bernadis, P. *et al. Nature* **404**, 955–959 (2000).
5. Hu, W., Sugiyama, N. & Silk, J. *Nature* **386**, 37–43 (1997).
6. Bahcall, N. *et al. Science* **284**, 1481–1488 (1999).
7. Fukugita, M., Hogan, C. J. & Peebles, P. J. E. *Astrophys. J.* **503**, 518–530 (1998).
8. Peebles, P. J. E. & Yu, J. T. *Astrophys. J.* **162**, 815–836 (1970).

Molecular biology

Introns gain ground

Thomas H. Eickbush

Protein-encoding genes of eukaryotes are often disrupted by introns that must be removed from the primary RNA transcript before the transcript is translated into protein. The origin of these apparently useless gene segments, known as spliceosomal introns because of the machinery used to excise them from RNA, has been debated for nearly 25 years. They were initially thought to be remnants of the original formation of genes. But there is mounting evidence that they represent the insertion of mobile elements from elsewhere, the most likely suspects being the mobile group II introns found in bacteria, mitochondria and chloroplast genomes. On page 1018 of this issue¹, Cousineau and co-workers provide further support for this possibility by revealing a means by which group II introns can spread throughout a genome.

A link between group II introns and spliceosomal introns was initially suggested by the remarkable similarity of the two RNA-splicing reactions². Both intron types splice through a two-step transesterification mechanism. The first step is the formation of an intron 'lariat' in which the nucleotide at the 5' end of the intron is covalently linked to an adenine near its 3' end. This is followed by joining of the protein-coding exons, coupled to cleavage at the 3' splice site and release of the intron lariat. In addition to this similarity in chemistry, the sequences at the 5' and 3' junctions of the two types of introns are quite similar.

The main difference between the splicing reactions is that the group II introns are self-splicing — they contain highly conserved internal structures that catalyse their own splicing reactions. Removal of nuclear spliceosomal introns, on the other hand, is catalysed by the spliceosome, an elaborate external assembly of small nuclear RNAs and associated proteins. A second feature associated only with group II introns is that some

are mobile. So if an intronless copy of a gene is introduced into a cell with one of these mobile introns, a copy of the intron will insert into the equivalent intronless site, a process called homing³.

The first clue to the mechanism of this movement came with the discovery that the mobile introns encode a reverse transcriptase — an enzyme that can make a double-stranded DNA copy from a single-stranded RNA template. The detailed mechanism of group II intron 'retrohoming' was determined in an elegant combination of studies of introns from yeast mitochondria⁴ and of the LtrB intron of the bacterium *Lactococcus lactis*^{5,6}. As shown in Fig. 1a, the spliced intron lariat reverse splices into the DNA sense strand at the target site in the intronless gene. The antisense strand is cleaved by an intron-encoded protein with DNA endonuclease activity. Primary recognition of the DNA target site is by base pairing of two regions of the intron RNA and the exon sequences upstream of the insertion site. Synthesis of complementary DNA by the intron-encoded reverse transcriptase, and synthesis of the second DNA strand, complete the reaction.

As well as extensive pairing between the intron RNA and upstream exon sequences, efficient group II retrohoming requires endonuclease recognition of specific bases in the target DNA⁷. This would seemingly preclude group II introns as an origin of spliceosomal introns because a proliferation phase would be required to spread the introns to countless unrelated sites throughout the eukaryotic genome. However, Cousineau *et al.*¹ now show for the first time the ease with which group II introns can expand from their original homes and invade chromosomal sites elsewhere.

The LtrB group II intron of *L. lactis* used in these studies was marked in two ways — with an antibiotic-resistance gene to enable

introns to be selected and their movement followed, and with another marker that allowed Cousineau *et al.* to see whether mobility involved an RNA intermediate. The group II intron was maintained on a temperature-sensitive plasmid so that removal of the plasmid enabled the selection of cells with LtrB insertions in new chromosomal sites. Not only were several such events detected, but the LtrB insertions were all into the sense strand of various genes; moreover, it seems that all insertions occurred through an RNA intermediate. Remarkably, although the new sites had only limited sequence similarity to the original target site, the LtrB introns were spliced out of the RNA transcript from the newly inserted gene (albeit at low frequency). The mechanism of these

jumps was unlike that of retrohoming, as it did not require intron endonuclease activity but was dependent on the host's genetic recombination machinery.

The proposed model for these retrotranspositions to different chromosome sites is shown in Fig. 1b. Rather than targeting DNA, the LtrB intron is thought to reverse splice into an RNA transcript of another gene. The composite RNA is reverse transcribed into complementary DNA, and the second strand of DNA is synthesized. The host's recombination machinery then inserts the intron-bearing sequence into the host chromosome at a place that depends on the composition of the flanking sequences. Amazingly, this mechanism is precisely the one proposed for the spread of self-splicing introns soon after they were discovered^{8,9}.

The clear advantage of group II intron spreading by reverse splicing into RNA, rather than DNA, is that an endonuclease is not needed and that insertion into the sense strand of a gene is assured. Because the ability of an intron to reverse splice into new RNA suggests that it can likewise be spliced out, continued expression of the newly interrupted gene is highly likely. Once inserted, mutations that increased splicing efficiency would be selected for. At the same time, mutations in the endonuclease that permit recognition of the fresh DNA target site would enable retrohoming to occur, and so allow the rapid spread of the intron to this new site in other members of the species. As intron number increases, so does the chance of another retrotransposition to yet another new site; and so the cycle repeats.

Why did the descendants of group II introns lose their ability to self-splice? It could be that, as eukaryotic genes accumulated many introns, the cells evolved the machinery to increase the efficiency of the splicing reactions. In light of Cousineau and colleagues' discovery of group II intron jumping, it could also be that the cell's control over splicing was a necessary defence against these invasive elements. By assuming the catalytic activity, the cell could prevent them from spreading.

Are there other remnants of group II introns in the eukaryotic genome? Prevented from moving to new sites by reverse splicing, all that the reverse transcriptase and endonuclease of a group II intron needed to retain mobility was a means to cleave both strands of a DNA target. Indeed, forward splicing was now unnecessary as there were now many sites in the emerging eukaryotic genomes where additional insertions would have minimal consequences (newly existing introns, for example). Non-long-terminal-repeat retrotransposable elements are one of the most widespread and abundant mobile elements in modern eukaryotic genomes. Their reverse transcriptase is closely related to that of group II introns; and they use a



100 YEARS AGO

The Athenæum makes the following announcement:— The "Diary of White of Selborne" is to be published. He kept it, as is well known, for more than twenty-five years, and used for the purpose a form "invented" by Daines Barrington, entitled "The Naturalists' Calendar", constructed for recording on each day, in proper columns, the readings of the thermometer and barometer; the direction of the wind; the measurement of the rainfall; the weather; the appearance of leaves and flowers of plants; the appearance or disappearance of birds and insects; observations with regard to fish and other animals; and miscellaneous observations. But Gilbert White enriched his "Calendar" with much other matter. There are not only numerous disquisitions on points of natural history, but notes of events of public interest and of personal or domestic concern. From *Nature* 26 April 1900.

50 YEARS AGO

Essentially the whole question of secrecy [centring on atomic energy and its use] is a particular aspect of the fragmentation of society, and it is idle to imagine that standards of professional conduct or loyalty, for example, will be raised if we fail to take account of the sanctions which really uphold our ideals of freedom, or the profound philosophy and the spiritual discipline in which such freedoms and rights flourish. To improve the administration of the measures required to ensure that within the defined areas only reliable and loyal persons are employed is not enough. We must look further to the underlying causes for such breaks with the tradition of science and professional loyalty. The real threat to national security to-day lies deeper, in those processes in society which are eroding or repudiating the basis both of a free society and of conditions in which science itself flourishes. To some extent, even among men of science, confidence in the autonomy and invulnerability of science has been shaken since Mr. Middleton Murry's "The Free Society" was published two years ago, with its warning that science is as vulnerable as the free society itself, and is ultimately dependent on it. That we should still find men of science disregarding and even whittling down the continuous tradition of scientific integrity, created and continually renewed by conscience, is nevertheless a disturbing feature of the situation. From *Nature* 29 April 1950.

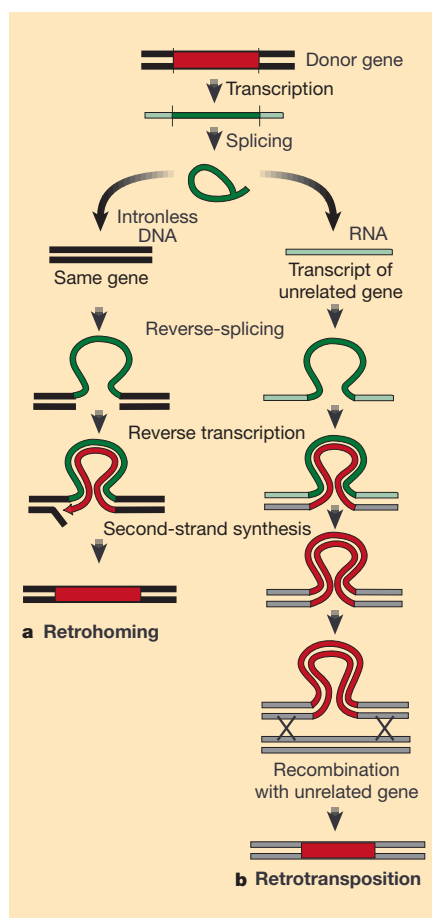


Figure 1 Retrohoming and retrotransposition models for group II introns. a, Retrohoming. The excised intron lariat can reverse splice with high efficiency into an intronless site in DNA of the same gene. Reverse transcription and synthesis of the second strand of DNA result in the intron's incorporation at the new site. b, Retrotransposition. Cousineau *et al.*¹ propose that the excised group II intron can also reverse splice into an RNA transcript of another gene, although at a much lower frequency than in retrohoming. Reverse transcription of this composite RNA and its recombination back into the new DNA site results in the spread of the intron.

reverse transcription mechanism, primed by target DNA, that is similar to group II retro-homing¹⁰. Perhaps these are the mobile descendants of group II introns. ■

Thomas H. Eickbush is in the Department of Biology, University of Rochester, Rochester, New York 14627, USA.

e-mail: eick@mail.rochester.edu

1. Cousineau, B., Lawrence, S., Smith, D. & Belfort, M. *Nature* **404**, 1018–1021 (2000).

2. Jacquier, A. *Trends Biochem. Sci.* **15**, 351–354 (1990).
3. Michel, F. & Ferat, J. *Annu. Rev. Biochem.* **64**, 435–461 (1995).
4. Eskes, R., Yang, J., Lambowitz, A. M. & Perlman, P. S. *Cell* **88**, 865–874 (1997).
5. Matsuura, M. *et al. Genes Dev.* **11**, 2910–2924 (1997).
6. Cousineau, B. *et al. Cell* **94**, 451–462 (1998).
7. Mohr, G., Smith, D., Belfort, M. & Lambowitz, A. M. *Genes Dev.* **14**, 559–573 (2000).
8. Sharp, P. A. *Cell* **42**, 397–400 (1985).
9. Cech, T. R. *Int. Rev. Cytol.* **93**, 3–22 (1985).
10. Yang, J., Malik, H. & Eickbush, T. H. *Proc. Natl Acad. Sci. USA* **96**, 7847–7852 (1999).

Nanotechnology

Heat flow through nanobridges

Leo P. Kouwenhoven and Liesbeth C. Venema

When a sample of material is heated on one side, energy will flow to the cold end. In solid material, heat can be carried either by mobile electrons or by atoms vibrating around their fixed, equilibrium positions. Insulating materials do not contain mobile electrons so only atomic vibrations can transport heat in insulators. These vibrations are not random: the atoms move collectively so that together they form waves, called phonons. Objects in everyday life, such as a ceramic cooking pot, contain many different types of phonons with varying wavelengths, which carry heat from the hot to the cold end. But by making objects smaller and smaller, today's trend in physics and engineering, properties of phonons emerge that are not evident in large materials. On page 974 of this issue, Schwab *et al.*¹ report an experiment involving an extremely narrow (nanometre-scale) bridge for phonons. A nanometre is a billionth of a metre, making the bridges about 500 atoms wide. The authors find that the amount of heat that can flow across the nanobridge is bound by an upper limit, set by the laws of quantum mechanics.

In the world of nanoscale objects, such as the nanoharp shown in Fig. 1a, heat cannot be carried by phonons of just any wavelength. Of all possible phonons, the one with the longest wavelength that just fits within the material has the lowest energy. Longer wavelengths (and hence lower-energy phonons) are not allowed. For smaller objects, the maximum wavelength decreases, and consequently the minimum phonon energy increases. When the object is so small that this minimum phonon energy exceeds the thermal energy, $k_B T$ (where k_B is the Boltzmann constant and T the temperature), quantum behaviour of the phonon motion can be expected.

Many types of phonons with different wavelengths and characters can exist. Each leads to a specific mode of wave-like motion in an object, such as twisting or bending in a wire. In the centre of the device created by

Schwab *et al.*¹, various phonons are generated by heating (Fig. 1b). But only specific phonon modes can couple into the narrow

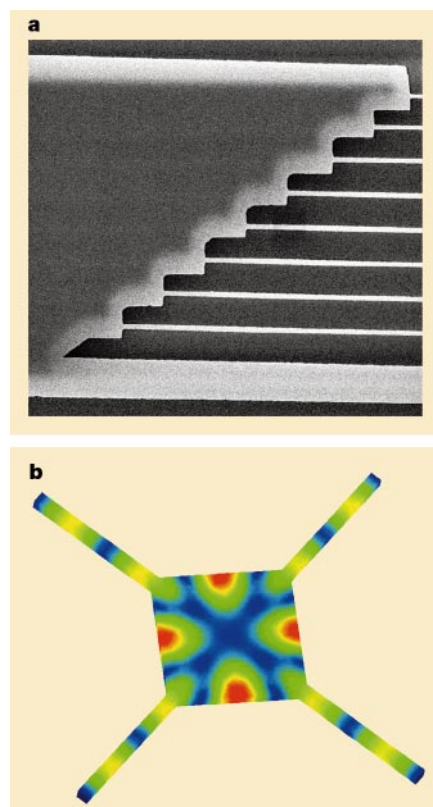


Figure 1 The nanoscale world. **a**, Parallel nanobridges forming a nanoharp. The sound in the bridges of the nanoharp is restricted to certain resonant frequencies, as in a full-sized harp. Both sound and heat are carried through solids by phonons. The word phonon refers to the Greek word 'phon' for sound. (Photo courtesy of H. G. Craighead, Cornell University.) **b**, In the device created by Schwab *et al.*¹, phonons are generated in the central square by thermal heating. Some of these phonons can leave this region by way of the four suspended nanobridges. The colours indicate the local strain amplitude. (Simulation by D. Harrington, Caltech.)

wires attached at each corner. One of these possible modes is shown in colour. Because only specific modes are allowed, the phonon spectrum is quantized in these nanobridges. Similarly, oscillations in the wires of the nanoharp (Fig. 1a) can occur only at a particular set of discrete frequencies, much as the strings in a full-sized harp are restricted to resonate at certain frequencies. In the nanoharp, these frequencies form the quantized spectrum of allowed wavelengths.

How does a quantized phonon spectrum affect the flow of heat? Dimensionality plays an important role here. Each nanobridge shown in Fig. 1a is a one-dimensional (1D) system, in which phonons are free to move only in the longitudinal direction. In general, 1D systems display many quantized transport phenomena, irrespective of what is being transported (electrons, photons or other carriers)². For instance, the flow of electrons through a narrow constriction between two large conductors leads to a quantized conductance in units of $2e^2/h$ (where h is the Planck constant and e is the charge on an electron). In other words, each mode of electrical transport contributes a maximum value (independent of the material involved) of $2e^2/h$ to the conductance.

Quantized electrical conductance has turned out to be a very general phenomenon. It was first discovered in sub-micrometre-scale semiconductor transistors^{3,4}. Later, quantized conductance was also found in systems of two metals connected by a single atom⁵ and subsequently in carbon nanotubes that are dipped in a conducting solution⁶. Besides electrons, analogous 1D quantum transport phenomena have been observed for photons, for pairs of electrons in superconductors, and for heat carried by electrons².

Phonon transport is predicted to show similar quantized behaviour. This may seem surprising, because an electric current transports charge and mass, whereas a heat current carried by phonons transports no mass. Moreover, electrons and phonons obey different quantum statistics. Nonetheless, one single formula describes it all: the Landauer formula treats conduction as a general transmission problem between two reservoirs and is applicable to both electrons and phonons⁷.

Notwithstanding the theoretical similarity, the experimental requirements are hugely different. Commercial batteries and amperemeters are sufficiently sensitive to measure quantized electrical conductance. (Note that $(2e^2/h)^{-1}$ is equal to 13 k Ω , and is therefore an easy resistance to measure.) In contrast, for the observation of quantized phonon transport between two heat reservoirs, various technical difficulties have to be solved. First, to achieve good thermal isolation of the phonon wires they have to be freely suspended (Fig. 1a). This is not an easy

task on the scale of a hundred nanometres. In addition, a sensitive technique is needed to heat one end of the wire and measure the temperature difference between the two ends with millikelvin resolution. Schwab *et al.*¹ have managed to construct such a nanobridge device and have performed measurements at ultra-low-power levels. They find that a single phonon mode can at most contribute a quantum of $\pi^2 k_B^2 T/3h$ to the thermal conductance.

In experiments with quantized electron transport, a sequence of conductance steps is observed as the electronic states fill up. With phonons, in contrast, the authors find a single plateau at low temperatures, which then rises linearly with temperature at higher temperatures. The reason is that electronic states are either full or empty — the change in occupation occurs in a sharp, discontinuous step. The occupation of phonon states is actually regulated by the temperature (as indicated by the linear dependence of $\pi^2 k_B^2 T/3h$ on T). When the temperature is increased, more phonon states become occupied. But instead of a series of steps,

the hallmark of 1D quantum transport of phonons is saturation of the thermal conductance at very low temperatures, precisely at the quantum value $\pi^2 k_B^2 T/3h$.

This observation by Schwab *et al.* is the first demonstration of quantum physics in nanomechanical structures. We can now look forward to measurements of phonon counting, double-slit interference experiments with single phonons, and many other experiments on phonons in quantum systems. ■

Leo P. Kouwenhoven and Liesbeth C. Venema are in the Department of Applied Physics, Delft University of Technology, PO Box 5046, 2600 GA Delft, The Netherlands.

e-mails: leo@qt.tn.tudelft.nl

venema@qt.tn.tudelft.nl

1. Schwab, K., Henriksen, E. A., Worlock, J. M. & Roukes, M. L. *Nature* **404**, 974–977 (2000).
2. Van Houten, H. & Beenakker, C. W. J. *Phys. Today* 22–27 July (1996).
3. Van Wees, B. J. *et al.* *Phys. Rev. Lett.* **60**, 848–850 (1988).
4. Wharam, D. A. *et al.* *J. Phys. C* **21**, 209–214 (1988).
5. Muller, C. J., van Ruitenbeek, J. M. & de Jongh, L. J. *Phys. Rev. Lett.* **69**, 140–143 (1992).
6. Frank, S. *et al.* *Science* **280**, 1744–1746 (1998).
7. Imry, Y. & Landauer, R. *Rev. Mod. Phys.* **71**, S306–S312 (1999).

the striatum)⁴ and the deposition of polyglutamine aggregates both within and outside nuclei (primarily in neurons of the brain's cortex)^{5,6}. The precise relationship between the aggregation process and the malfunction and death of nerve cells is uncertain.

Mice that have been engineered to express exon 1 of the *HD* gene with long CAG repeats develop a phenotype with many features reminiscent of Huntington's disease⁷. In this model, a late-onset, progressively worsening movement disorder is preceded by the formation of the characteristic polyglutamine aggregates^{8,9}. At the same time that symptoms appear, the binding of a particular neurotransmitter (dopamine) to dopamine receptors on neurons decreases, as do levels of the messenger RNA encoding that receptor¹⁰. The brains of Huntington's patients show a similar pattern¹¹.

Yamamoto *et al.*¹ have built on this work, taking advantage of a conditional regulatory system¹² that allows the mutant exon 1 *HD* gene to be switched on or off. They found that, when the gene was switched on in mice, it was expressed throughout the animals' lifetime. The mice developed a progressive neurological phenotype: onset of a motor disorder was seen at 4 weeks of age; a mild tremor was present by 20 weeks; and mice were clearly underactive by 36 weeks. Yamamoto *et al.* detected nuclear staining specific to the mutant *HD* gene, as well as intranuclear and extranuclear polyglutamine aggregates, in brain sections of mice at 8 weeks. The brains of the mice were smaller than normal, with a marked reduction in the size of the striatum. A reduction in binding of dopamine to its receptors was also seen. This is all in line with what would

Huntington's disease

In reverse gear

Gillian Bates

Yamamoto and colleagues¹, writing in *Cell*, describe a mouse model of Huntington's disease with a difference: the pathological features and symptoms of the disease can be reversed. These findings are both dramatic and unexpected, and indicate — with the obvious caveat that we do not yet know how these results will translate to humans — that it may be possible to treat this devastating disease after symptoms appear.

Huntington's disease is an inherited, late-onset neurodegenerative disease for which there is no known cure. In the Western world it affects about 1 in 10,000 people, with a slightly larger number than this being at risk of developing the disease. Symptoms include changes in personality, cognitive decline and a movement disorder (such as uncontrollable, jerky movements, called chorea)².

Patients suffering from Huntington's disease bear alterations in the *HD* gene³. Close to the beginning of this gene — in a protein-coding part, exon 1 — is a region that includes a variable number of a trinucleotide motif (the cytosine/adenine/guanine, or CAG, motif). The only difference between the *HD* gene in unaffected and affected individuals is the number of copies of this motif. The CAG trinucleotide encodes the amino acid glutamine, and so the huntingtin protein encoded by the *HD* gene contains a

polyglutamine tract of varying length. When the tract is 38 glutamines long or more, the protein gains toxic properties. On autopsy, the brains of Huntington's patients show a selective death of neuronal cells (mainly in

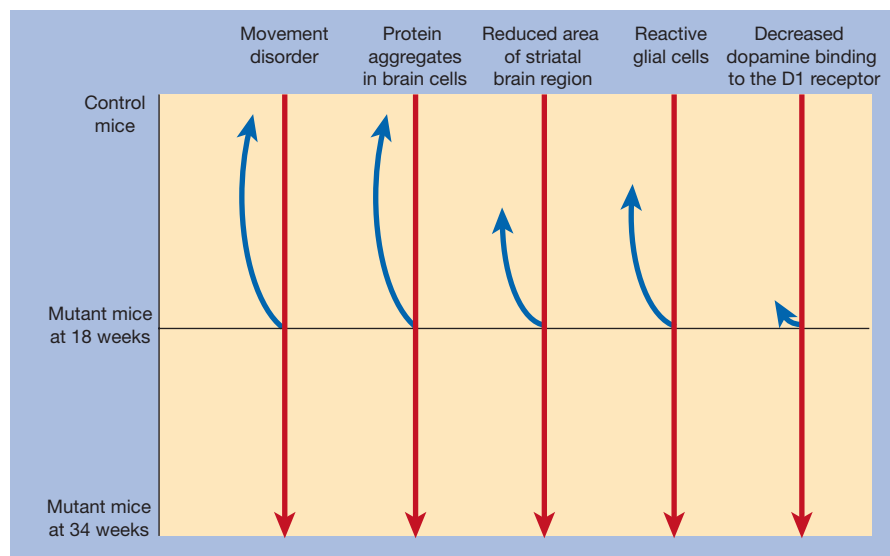


Figure 1 Progression and reversal of Huntington's disease in a mouse model. Red lines represent the relative progression of the various characteristics of Yamamoto *et al.*'s mutant mice¹ over a 34-week period. In one set of mice, the mutant gene was turned off at 18 weeks of age by adding the antibiotic doxycycline to the animals' drinking water. The blue lines represent the degree to which each of the characteristics shown was either halted or reversed when the gene was no longer expressed. Dopamine is a neurotransmitter; reactive glial cells are generally indicative of nerve cell loss.

be expected of a mouse model of Huntington's disease.

But the unexpected happened when the gene was switched off. Yamamoto *et al.* did this by administering an antibiotic in the animals' drinking water; this antibiotic inhibited an activator of the *HD* gene that was also inserted into the mice. In a group of the mice, the *HD* gene was kept switched on until the mice were 18 weeks old. By this point, the animals showed clear symptoms and neurological features of the disease. Then, half of the group were continually given the antibiotic ('gene-off' group), while the other half were maintained with the gene switched on ('gene-on' group). After another 16 weeks, in the gene-off group the nuclear staining and intranuclear and extranuclear polyglutamine aggregates had disappeared from the striatum, and were much reduced in the cortex (Fig. 1). These mice showed no further reduction in striatal size and the decrease in binding of dopamine to its receptors was slightly reversed, compared with the 18-week time point. But an equally remarkable finding was that the progression of the motor disorder had been reversed to levels approaching those of control animals.

This reversal of aggregate accumulation and neurological symptoms would have been difficult to predict. Polyglutamine aggregates are highly insoluble, and are difficult to dissolve or denature biochemically^{13,14}. But Yamamoto *et al.* have shown that the neuron is capable of dispensing with such structures. It seems likely that aggregates can be dismantled and made accessible to the proteasome — the cell's main protein-degrading machinery — and so can be destroyed by common cellular mechanisms.

One open question in this area of research is whether or not the protein aggregates are responsible for the onset and progression of the disease. Although Yamamoto *et al.*¹ have shown that aggregation and symptoms are reversed in parallel in their mice, they cannot establish a definitive, causal link between the aggregation pathway and the disease. But disease progression clearly depends on the continuous expression of the mutant *HD* gene.

What does all this mean for our ability to apply the brakes to Huntington's disease in humans? Turning off the mutant *HD* gene (and hence production of the mutant huntingtin protein) reverses the disease symptoms in the mice. So it seems that neither the soluble nor the aggregated form of the protein irreversibly commits a neuron to die. At least the early symptoms are probably caused not by neuronal death, but rather by a neuronal malfunction that can be reversed. So, therapeutic approaches to reduce the expression of the mutant protein or to block any of the steps that lead to the formation of aggregates might not only prevent or delay the onset of the disease. They may also offer a

cure — if used when neuronal malfunction can still be turned around.

Gillian Bates is in the Division of Medical and Molecular Genetics, GKT School of Medicine, King's College, 8th Floor Guy's Tower, Guy's Hospital, London SE1 9RT, UK.

e-mail: gillian.bates@kcl.ac.uk

1. Yamamoto, A., Lucas, J. J. & Hen, R. *Cell* **101**, 57–66 (2000).
2. Harper, P. S. *Huntington's Disease* (Saunders, London, 1996).
3. Huntington's Disease Collaborative Research Group *Cell* **72**, 971–983 (1993).
4. Vonsattel, J.-P. *et al.* *J. Neuropathol. Exp. Neurol.* **44**, 559–577 (1985).

5. DiFiglia, M. *et al.* *Science* **277**, 1990–1993 (1997).
6. Gutekunst, C.-A. *et al.* *J. Neurosci.* **19**, 2522–2534 (1999).
7. Mangiarini, L. *et al.* *Cell* **87**, 493–506 (1996).
8. Davies, S. W. *et al.* *Cell* **90**, 537–548 (1997).
9. Li, H. *et al.* *Hum. Mol. Genet.* **8**, 1227–1236 (1999).
10. Cha, J.-H. *et al.* *Phil. Trans. R. Soc. Lond. B* **354**, 981–989 (1999).
11. Augood, S. J., Faull, R. L. & Emson P. C. *Ann. Neurol.* **42**, 215–221 (1997).
12. Gossen, M. & Bujard, H. *Proc. Natl Acad. Sci. USA* **89**, 5547–5551 (1992).
13. Scherzinger, E. *et al.* *Cell* **90**, 549–558 (1997).
14. Kasantsev, A. *et al.* *Proc. Natl Acad. Sci. USA* **96**, 11404–11409 (1999).

Signal transduction

Molecular switches in lipid rafts

Leslie A. Cary and Jonathan A. Cooper

Cells do not exist in isolation from their external environment, and need to be able to modify their activity according to what is happening outside them. Molecular switches act to relay information from a cell's exterior to its interior. One set of molecular triggers are the Src-family kinases (SFKs), the best known of which, Src, lends its name to the family. The SFKs regulate a plethora of events, including cell growth, division, differentiation, survival and death, as well as specialized functions such as immune responses, the attachment of cells to extracellular matrix, cell movement, and uptake of molecules from the cell surface^{1,2}. Unrestricted SFK activity can lead to cancer, but ordinarily SFKs are held in a stable, inactive state until needed.

The switch between the inactive and active states of an SFK involves both conformational changes and modification by phosphorylation — the addition of phosphate groups to particular amino-acid residues of the protein. But it is still unclear how the cell coordinates these events in response to extracellular stimuli¹. A significant step forward has now been taken by Kawabuchi and colleagues³. Writing on page 999 of this issue, they identify a new protein, called Cbp, that links SFKs to a negative regulator.

The SFKs are a subset of enzymes known as tyrosine kinases, which transfer phosphate groups onto the amino-acid residue tyrosine in proteins. A characteristic of a phosphorylated tyrosine residue is its ability to bind a protein region known as an SH2 domain. Non-phosphorylated tyrosine residues, in contrast, cannot bind SH2 domains. So, by phosphorylating specific residues, tyrosine kinases can mediate the binding of one protein to another, or induce conformational changes within proteins that modify their activity.

Many tyrosine kinases span the membrane encapsulating a cell, and are activated only by binding of one specific extracellular molecule. In contrast, SFKs associate with

the cell membrane but reside totally inside the cell, and are activated in response to a variety of extracellular stimuli. An SFK is in an inactive state when it is itself phosphorylated at an inhibitory carboxy-terminal tyrosine; the protein then folds up into an inactive conformation in which its inhibitory tyrosine residue binds to the protein's own SH2 domain (Fig. 1a, page 947). The SFK is activated when this phosphate is removed by a phosphatase, or when another protein binds and displaces these intramolecular interactions. These events result in a different tyrosine residue becoming exposed and phosphorylated, thus activating the protein. Once active, the SFK can phosphorylate other proteins and then bind to them through its now exposed SH2 domain.

Phosphorylation of the inhibitory tyrosine is done by another kinase, known as Csk (for carboxy-terminal Src kinase). And, at least in lymphocytes (immune cells), Csk has a second way of inactivating SFKs. In these cells, Csk is bound to a phosphatase, called PEP, that seems to specifically dephosphorylate the second (activating) phosphorylation site in the SFK⁴. So, the Csk–PEP complex both adds an inhibitory phosphate to SFKs and removes an activating phosphate.

Csk (and a closely related kinase of similar function, Chk) is thus critical for keeping SFKs inactive. Indeed, removal of the Csk gene deregulates SFK activity^{5–7}. But is Csk itself regulated in some way? Or is it always 'on', ready to inactivate SFKs soon after they have been activated? Unlike SFKs, Csk is not tyrosine-phosphorylated at either inhibitory or stimulatory sites. However, it does contain an SH2 domain, and Kawabuchi *et al.*³ now provide compelling evidence that Csk uses this domain to relocate from the cytosol to the vicinity of active SFKs (that is, to the membrane). So, Csk might be regulated spatially more than catalytically.

Kawabuchi *et al.* find that Csk uses its SH2 domain to bind to a new protein, Csk-binding protein (Cbp), when the latter is

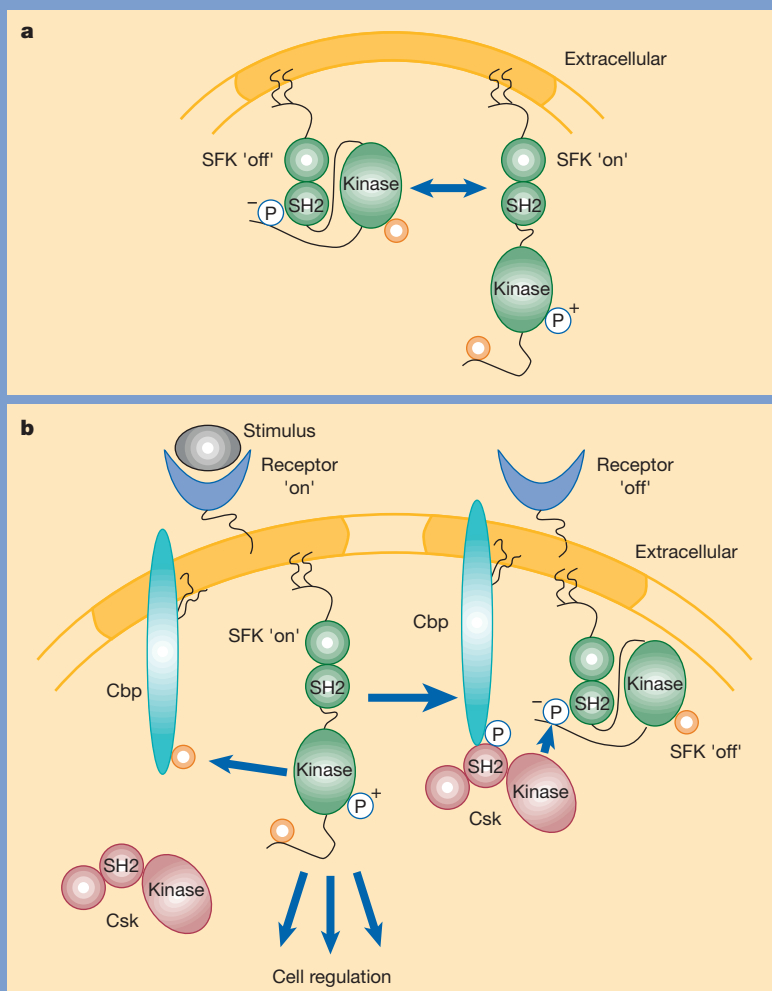


Figure 1 Turning a molecular switch on and off. Src-family kinases (SFks) are molecular switches that relay extracellular information to the cell's interior. **a**, Left, when phosphorylated (circled 'P') at its inhibitory tyrosine residue (minus symbol), an SFK forms an inactive conformation as a result of binding of this tyrosine to the SFK's SH2 domain. Right, dephosphorylation of this residue (empty orange circle) induces an activated conformation, with accessible kinase and SH2 domains and a phosphorylated stimulatory tyrosine residue (plus symbol). **b**, As shown by Kawabuchi *et al.*³, an SFK might be localized with Cbp and cell-surface receptors in membrane regions called 'rafts' (dark yellow areas). Left, binding of an extracellular stimulus to its cell-surface receptor leads to the activation of an SFK, which then phosphorylates other proteins to regulate cellular function and might also phosphorylate a tyrosine residue of Cbp. Right, this phosphorylated tyrosine allows the binding of Cbp to the SH2 domain of Csk, and so recruits Csk to the membrane. Csk can now inactivate the SFK by phosphorylating the SFK's inhibitory tyrosine residue.

phosphorylated at a specific tyrosine residue. Although Csk floats free in the cytosol, Cbp is firmly anchored in the membrane. So, binding to phosphorylated Cbp brings Csk to the membrane, where the SFks are found. Cbp is phosphorylated not by Csk but apparently by another tyrosine kinase — quite possibly an SFK. From these results, we get to the model shown in Fig. 1b. A cell-surface receptor binds to its particular extracellular partner and activates a specific SFK. The active SFK phosphorylates many proteins of varying cellular functions, and possibly Cbp as well. Cbp can now recruit Csk to the membrane, where it can phosphorylate

and inactivate the SFK, turning off the signalling events at the appropriate time.

The discovery of Cbp explains several observations. First, we knew previously that Csk's SH2 domain is important for SFK regulation, but we did not know why^{8,9}. Second, we knew that Csk relocates to the membrane — specifically to regions at which SFks are active — but we did not know how^{8,9}. Third, genetic analysis indicates that SFks are activated by certain phosphatases^{10,11}. However, these phosphatases do not show the expected specificity towards the inhibitory tyrosine on SFks. Kawabuchi *et al.*³ suggest that these phosphatases might instead act mainly on

Cbp, so promoting the activation of SFks indirectly.

Most SFks are modified with specific lipids that bring them to subdomains of the cell membrane that have high cholesterol and glycolipid content, called membrane 'rafts'¹ (Fig. 1). A number of factors that activate SFks bind to stimulus receptors that do not penetrate through the membrane but are instead concentrated in the rafts in the outer membrane, pointing outwards from the cell surface¹². Interestingly, Cbp is also found in rafts — probably by virtue of the same lipid modification as the SFks³. Even though Kawabuchi *et al.* could not detect a complex between Src and Cbp, it is likely that most SFks would be near Cbp in rafts. So, we speculate that one way in which outer-membrane receptors located in rafts could stimulate SFks is if they were to alter the balance between phosphorylation and dephosphorylation of Cbp. The fact that Cbp has extracellular and transmembrane regions might allow it to interact with the receptor.

Coincidentally, Brdička *et al.*¹³ have independently identified Cbp (which they call PAG) as an SFK-associated protein in rafts purified from lymphocytes. They show that Cbp can be phosphorylated by SFks and associates with Csk, and also that Cbp becomes dephosphorylated when lymphocytes are stimulated, at the same time that SFks are activated. So, dephosphorylation of Cbp coincides with, and probably contributes to, SFK activation.

Thus, two groups^{3,13} have identified what could be a key regulator of SFks. If they are right, ablation of the Cbp gene might have similar effects to the loss of Csk — that is, constitutive SFK activation. If so, turning off Cbp expression would be one way for cancer to arise, and the Cbp gene might be lost during the evolution of human cancers. A lot more should be learnt from experiments in which the Cbp gene is knocked out — experiments that are eagerly awaited.

Leslie A. Cary and Jonathan A. Cooper are at the Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, PO Box 19024, Seattle, Washington 98109-1024, USA. e-mail: jcooper@fred.hcr.org

1. Thomas, S. M. & Brugge, J. S. *Annu. Rev. Cell Dev. Biol.* **13**, 513–609 (1997).
2. Wilde, A. *et al. Cell* **96**, 677–687 (1999).
3. Kawabuchi, M. *et al. Nature* **404**, 999–1003 (2000).
4. Cloutier, J. F. & Veillette, A. *J. Exp. Med.* **189**, 111–121 (1999).
5. Imamoto, A. & Soriano, P. *Cell* **73**, 1117–1124 (1993).
6. Nada, S. *et al. Cell* **73**, 1125–1135 (1993).
7. Schmedt, C. *et al. Nature* **394**, 901–904 (1998).
8. Sabe, H., Hata, A., Okada, M., Nakagawa, H. & Hanafusa, H. *Proc. Natl Acad. Sci. USA* **91**, 3984–3988 (1994).
9. Howell, B. W. & Cooper, J. A. *Mol. Cell. Biol.* **14**, 5402–5411 (1994).
10. Ponniah, S., Wang, D. Z., Lim, K. L. & Pallen, C. J. *Curr. Biol.* **9**, 535–538 (1999).
11. Thomas, M. L. & Brown, E. J. *Immunol. Today* **20**, 406–411 (1999).
12. Stefanova, I., Horejsi, V., Ansoltegui, I. J., Knapp, W. & Stockinger, H. *Science* **254**, 1016–1019 (1991).
13. Brdička, T. *et al. J. Exp. Med.* (in the press).

Superconducting nanowires

Gerd Schön

The continuing miniaturization of electronic devices has reached system sizes of ten nanometres, where quantum mechanical effects become important. A fundamental question is how the collective properties of these devices, such as superconductivity, are affected when the system's dimensions are reduced. Another issue is how energy dissipation influences quantum effects, eventually leading to a crossover to classical behaviour.

On page 971 of this issue, A. Bezryadin, C. Lau and M. Tinkham¹ report the results of experiments with ultrathin superconducting wires fabricated using an original technique. They show that a thin superconducting wire is a quantum system, which can be superconducting or insulating depending on the interaction of the system with its environment. The study of these non-classical effects is crucial in order to establish the fundamental limits of conventional devices, and because such effects open new perspectives for future applications, including 'quantum-state engineering' based on superconducting elements^{2,3}.

Quantum effects have been studied for many years in semiconductor systems. In quantum dots, in which electron gases are confined to two dimensions, discrete energy levels have been observed by optical and tunnelling spectroscopy (for review see ref. 4). Various techniques have been used to fabricate ultrathin wires and point contacts for quantum circuits, in which only a few conducting channels determine electron transport. In metallic systems, owing to the higher electron density and hence shorter electron wavelength, quantum effects are less prominent. It only recently became possible to observe discrete energy levels in the current-voltage characteristics of tunnelling electrons in metallic grains. It is understood that once the energy-level spacing becomes comparable to the superconducting energy gap of the bulk material, the superconductivity is destroyed⁵.

Since the early days of superconductivity it has been clear that thin superconducting wires may actually — because of phase-slip processes — exhibit resistive behaviour. The superconducting state can be characterized by an 'order parameter' — in effect, a macroscopic wavefunction describing the behaviour of the superconducting electrons, which form Cooper pairs. Superconductivity depends on the long-range phase coherence of the wavefunction: when the order parameter is non-zero and the phase is coherent, the resistance vanishes, and current

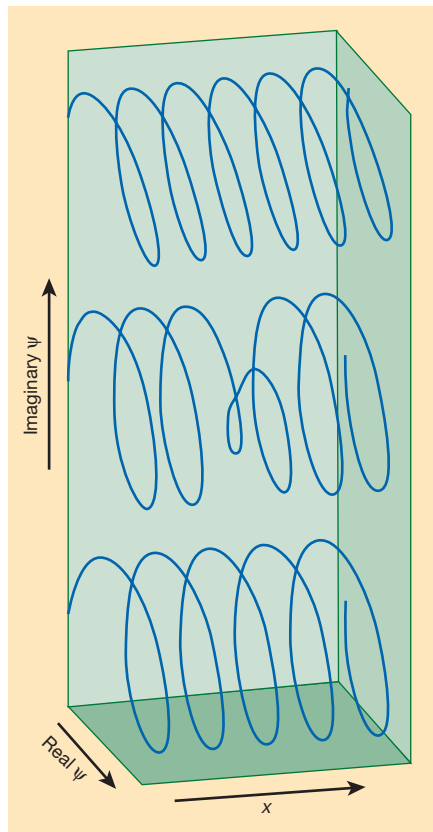


Figure 1 A phase-slip process. Superconductivity depends on the phase coherence of the complex order parameter, ψ , shown here. During a phase slip, the phase of the order parameter is reduced by 2π . The top image shows the order parameter before the phase slip, the middle image at some instant during the process, and the bottom image after the process. Phase-slip processes are responsible for the resistance of superconducting wires. At very low temperatures a quantum phase transition separates regimes where they are rare or frequent, corresponding to superconducting or insulating behaviour. Bezryadin *et al.*¹ observe such a transition in ultrathin superconducting wires.

flows without dissipation. In a phase-slip process, the order parameter shrinks, for a short timespan and over a narrow region, to zero, allowing its phase to slip by 2π , before it recovers to a normal value (Fig. 1).

Each phase slip generates a voltage pulse on the wire, resulting in dissipation of energy and increasing resistance. (Dissipated power is simply current times voltage.) Near the superconducting transition temperature, T_c , these phase slips occur frequently, so the resistance of a wire decreases only gradually as the temperature is lowered

below T_c . Langer and Ambegaokar developed the first theoretical description of these phenomena in the late 1960s. In their picture the resistance of the superconductor is governed by thermally activated phase slips, in which the system makes a transition across a potential barrier between two different metastable states (local minima corresponding to the top and bottom images in Fig. 1). Experiments have confirmed this picture and the results are well documented in textbooks⁶.

At temperatures close to absolute zero, where the thermal energy $k_B T$ is lower than the characteristic energy of the phase-slip process, $\hbar\omega$, quantum effects become relevant. So the resistance of the superconductor is caused by phase-slip processes that involve quantum tunnelling through the potential barrier⁷, rather than thermal activation over the barrier. Estimates show that the quantum phase-slip rate is usually too low to be observable in an experiment, unless the wire is made extremely narrow. After earlier claims, which later turned out to be noise effects, Giordano⁸ found experimental evidence for this quantum process. But his work was criticized, in part because the influence of dissipation had not been studied.

The experiments of Bezryadin *et al.*¹ are convincing because of their sample preparation as well as their analysis. They fabricate homogeneous superconducting wires less than 10 nm in width, which is below what can be achieved by conventional electron-beam lithography. To do so they started from probably the best available one-dimensional system, a carbon nanotube, which they coated with a superconducting material. A similar idea was pursued with some success by Braun *et al.*⁹, who managed to coat DNA with metal, but produced only rather disordered chains of metal clusters. In contrast, Bezryadin *et al.* managed to produce homogeneous superconducting wires.

They find that ultrathin wires become superconducting only as the temperature is lowered if their normal-state resistance is lower than a critical resistance, whereas samples with higher normal-state resistance remain resistive (even insulating). This behaviour can be interpreted as a 'quantum phase transition', in which the quantum mechanical behaviour of the system changes when the critical point is crossed. Here, the critical value of the resistance at which the behaviour changes corresponds to the quantum resistance for Cooper pairs ($R_q = 6.45 \text{ k}\Omega$).

Theoretically, several quantum phase transitions between superconducting and insulating behaviour have been predicted for superconducting systems. A single Josephson junction with resistive dissipation R should also show a transition at $R = R_q$ (ref. 10). Only last year Penttilä *et al.*¹¹

found indications of this transition. A similar transition has been predicted for dissipative as well as non-dissipative Josephson junction arrays. In the latter case it follows from the duality between charges and vortices^{12,13} and has been confirmed by experiments¹⁴.

Of greater relevance here, Zaikin *et al.*¹⁵ predicted a quantum phase transition in one-dimensional superconducting wires, which depends on the strength of the dissipation. The conclusion of this theory is that the low-temperature state of a thin wire actually has many phase-slip processes occurring at random times and positions. For finite-length wires ($\leq 1 \mu\text{m}$), Zaikin *et al.* predicted a quantum phase transition of the system between a superconducting state with 'bound' quantum phase slips and an insulating state with 'dissociated' phase slips at a critical value of the resistance R_q .

The experiments of Bezryadin *et al.*¹ agree remarkably well with these predictions¹⁵, confirming our picture of the way superconductivity vanishes in ultrathin wires. This work is crucial for developments in nanoelectronics and quantum-state engineering. Further experimental and theoretical studies will reveal more about transitions from classical to quantum mechanical behaviour in ultrasmall systems. The transition studied by Bezryadin *et al.*

depends strongly on the interaction of the quantum system with its environment (that is, dissipation). This may open the possibility of controlling the state — superconducting or insulating — of a nanowire by changing its environment.

Gerd Schön is at the Institut für Theoretische Festkörperphysik, Universität Karlsruhe, D-76128 Karlsruhe, Germany, and the Forschungszentrum Karlsruhe, Institut für Nanotechnologie, D-76021 Karlsruhe, Germany.

e-mail: schoen@tfp.physik.uni-karlsruhe.de

1. Bezryadin, A., Lau, C. N. & Tinkham, M. *Nature* **404**, 971–974 (2000).
2. Makhlin, Yu., Schön, G. & Shnirman, A. *Nature* **398**, 305–307 (1999).
3. Nakamura, Y., Pashkin, Yu. A. & Tsai, J. S. *Nature* **398**, 786–788 (1999).
4. Kouwenhoven, L. P. *et al.* in *Mesoscopic Electron Transport* (eds Sohn, L. L. *et al.*) NATO ASI Series E, Vol. 345, 105–214 (Kluwer, Dordrecht, 1997).
5. Ralph, D. C., Black, C. T. & Tinkham, M. *Phys. Rev. Lett.* **78**, 4087–4090 (1997).
6. Tinkham, M. *Introduction to Superconductivity* (MacGraw Hill, New York, 1996).
7. Mooij, J. E. & Schön, G. *Phys. Rev. Lett.* **55**, 114–117 (1985).
8. Giordano, N. *Phys. Rev. Lett.* **61**, 2137–2140 (1988).
9. Braun, E., Eichen, Y., Sivan, U. & Ben-Yoseph, G. *Nature* **391**, 775–778 (1998).
10. Schmid, A. *Phys. Rev. Lett.* **51**, 1506–1509 (1983).
11. Penttilä, J. S. *et al.* *Phys. Rev. Lett.* **82**, 1004–1007 (1999).
12. Fisher, M. P. A., Weichman, P. B., Grinstein, G. & Fisher, D. S. *Phys. Rev. B* **40**, 546–570 (1989).
13. Fazio, R. & Schön, G. *Phys. Rev. B* **43**, 5307–5320 (1991).
14. van der Zant, H. S. J., Geerligs, L. J. & Mooij, J. E. *Europhys. Lett.* **19**, 541–546 (1992).
15. Zaikin, A. D., Golubev, D. S., van Otterlo, A. & Zimanyi, G. T. *Phys. Rev. Lett.* **78**, 1552–1555 (1997).

Autoimmunity

Decoy receptors thwart B cells

Carl F. Ware

Lupus erythematosus is a debilitating autoimmune disease. It is characterized by pathological defects in many organs; by hyperactive immune cells called B lymphocytes, secreting self-reactive antibodies; and by a many-gene pattern of inheritance. A new piece to the physiological puzzle of this disease was revealed recently in the shape of a protein — called BAFF — that sends B cells into hyperdrive, leading to a lupus-like syndrome in mice^{1,2}. On page 995 of this issue³, Gross and colleagues define two receptors for this protein. They then use this key insight to construct a specific antagonist for BAFF that can slow the progression of certain features of the disease in autoimmune-prone mice. This work may have opened up a new avenue to the treatment of lupus erythematosus.

Many factors, both environmental and genetic — a lot of which remain ill defined⁴ — contribute to the onset of lupus erythematosus. Animal models have been instrumental in providing clues to the pathology of autoimmune diseases in humans. For example, the first generation of offspring of

the 'New Zealand B and W' mouse strains develop a lupus-like syndrome at around eight months of age: they have antibodies that react to their own nuclear molecules, including DNA, and a progressively worsening kidney disease⁵. The symptoms result in part from the immune-reactive complexes deposited in the networks of blood capillaries in the kidneys, and from the presence of activated B cells, which contribute to the inflammatory milieu. Eventually, kidney failure results — much as in the human disease.

Another mouse strain that is a good model of human autoimmunity — but not an exact model of lupus erythematosus⁶ — is the *lpr* (for lymphoproliferative disorder) strain. Mutations in a protein called Fas and its binding partner, both members of the tumour-necrosis factor (TNF) superfamily of proteins, cause the systemic autoimmune disease seen in these animals.

A pearl gleaned from database searching, BAFF was identified as a TNF homologue by several groups (and consequently labours under several different names, including zTNF4, as used by Gross *et al.*)^{2,4,7,8}. In tissue-

culture models, activated T lymphocytes, as well as immune cells called dendritic cells, express BAFF. When activated by antigen, these cells specifically enhance the proliferation of B lymphocytes^{2,7,9} — a key feature of autoimmune disorders.

Analysis of mice engineered to over-express the BAFF gene^{1–3} has established a further link between BAFF and autoimmunity, particularly lupus erythematosus. These transgenic animals have enlarged spleens and lymph nodes, which are filled with antigen-responsive, activated 'mature' B cells. By contrast, the number of T cells in these mice is slightly lower than normal, although some mice have a higher percentage of 'effector' T cells (those that control the spread of infections). The mice also have antibodies to DNA — a hallmark of lupus erythematosus — as well as other autoimmune antibodies. And, as the animals age, they show overt signs of kidney inflammation and subsequent loss of kidney function.

BAFF — like TNF itself — is a protein that spans the cellular plasma membrane. Both BAFF and TNF can be shed, in the shape of biologically active molecules, from the cell surface as a result of enzymatic action. The shed molecules circulate through the bloodstream and stimulate responsive B cells in distant tissues⁹ (Fig. 1). Gross *et al.*³ have

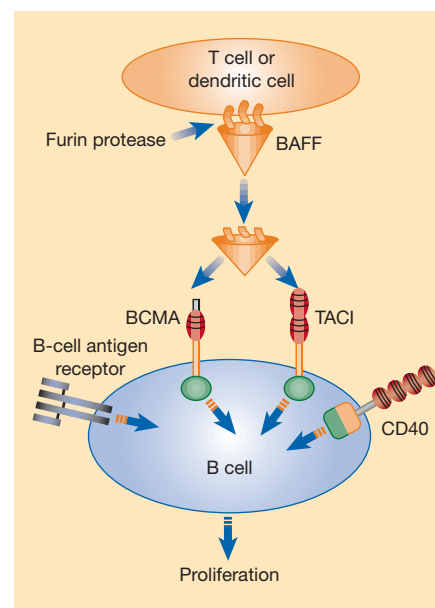


Figure 1 Lupus erythematosus in a mouse model. Gross *et al.*³ have defined a new pathway that stimulates B-cell proliferation and leads to autoimmunity in mice. BAFF protein is produced by activated T cells or dendritic cells. It is cleaved by furin protease, and the product is shed into the bloodstream. Gross *et al.* have shown that BAFF then interacts with BCMA or TACI, receptor proteins, on the surface of B cells. The B-cell antigen receptor and the BAFF receptors cooperate to stimulate B cells. Other activation signals are provided through the protein CD40.

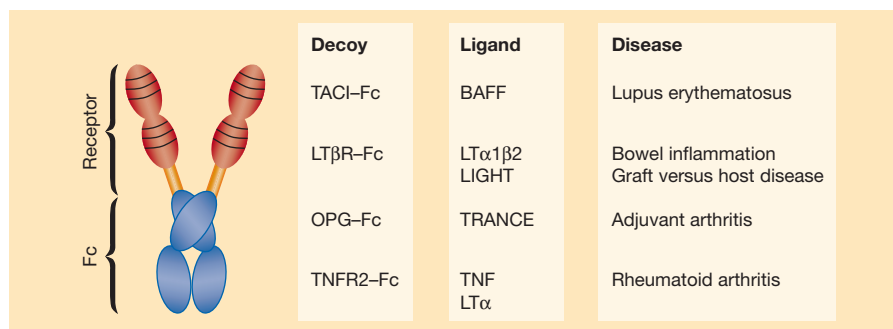


Figure 2 Decoy receptors suppress autoimmune diseases. Several members of the tumour-necrosis factor (TNF) family, including the lymphotoxin (LT)-β receptor (LTβR), osteoprotegerin (OPG) and the TNF receptors (including TNF receptor 2, or TNFR2) have significant roles in the pathogenesis of autoimmune disease, as shown by the ability of decoy receptors to inhibit symptoms. Gross *et al.*³ now add TACI to this list. The decoy receptors consist of part of the receptor in question, fused to the so-called Fc part of the antibody immunoglobulin G. Decoy receptors for TNF and lymphotoxin-α (LTα) have been successful in the treatment of rheumatoid arthritis. The other decoy receptors have shown an ability to ameliorate the symptoms of the diseases indicated in animal models.

extended these observations by showing that there is a strong correlation between the amount of circulating BAFF and disease progression in the serum of the New Zealand BW (first generation) and *lpr* strains of mice.

BAFF is clearly important in lupus erythematosus. All of these advances set the stage for treating the disease. But a key piece of the puzzle was missing — the receptor for BAFF on B cells.

Now, using an expression-cloning approach, Gross *et al.* have identified two genes that, when artificially introduced into baby hamster kidney cells, allow BAFF to bind to those cells. Sequencing of the genes showed that they encode two 'orphan' receptors — receptors without a known binding partner — in the TNF-receptor family. These receptors are called TACI and BCMA.

At first sight, these results might seem odd. Previous studies indicated that the BAFF receptor should be expressed only on B lymphocytes. But a study of TACI's history shows it to be linked to the activation of NF-AT — a transcription factor involved in activating T cells¹⁰. Likewise, the BCMA gene was identified in a T-cell lymphoma tumour¹¹ (the gene was fused to the interleukin-2 gene as a result of a chromosomal translocation). This is not at all what would be expected of the predicted BAFF receptor(s). However, further investigation revealed that these receptors are at least found in the right place: BCMA is expressed specifically in B cells, and TACI, although displayed on a subset of activated T cells, is also found on a subset of resting B cells¹⁰.

The next step for Gross *et al.* was to construct a decoy receptor for BAFF, by fusing the extracellular, ligand-binding domain of TACI or BCMA with a region of the antibody immunoglobulin G. As expected, the decoys bound to BAFF with high affinity and neutralized the stimulatory effect of BAFF on B-cell proliferation. And, when the TACI decoy was injected into first-generation mice

of the New Zealand BW strain, the amount of protein spillage into the urine was reduced, as was the number of B cells. All of this correlated with increased survival of the mice. Interestingly, however, Gross *et al.* saw no change in the amount of antibodies against DNA. This result reinforces findings that activated B cells contribute directly to disease progression in this model even when they do not produce their characteristic antibodies¹². The antibodies against DNA, although good markers of lupus erythematosus, may not actually cause the disease.

BAFF joins TNF and several related molecules as targets for decoy receptors that may prove useful in therapy for autoimmune diseases. Enthusiasm for these potential inhibitors is high, partly because of the success of decoy receptors for TNF and lymphotoxin in the clinical treatment of rheumatoid arthritis (Fig. 2). This enthusiasm should, of course, be tempered by the fact that lupus erythematosus involves many molecules and affects many tissues. History has taught us that targeting a single element is unlikely to produce the desired result. But BAFF and its receptors deserve serious attention as mediators of B-cell biology, with the promise of a better understanding of autoimmune diseases such as lupus erythematosus. ■

Carl F. Ware is in the Division of Molecular Immunology, La Jolla Institute for Allergy and Immunology, 10355 Science Center Drive, San Diego, California 92121, USA.
e-mail: carl_ware@liai.org

- MacKay, F. *et al.* *J. Exp. Med.* **190**, 1697–1710 (1999).
- Khare, S. D. *et al.* *Proc. Natl Acad. Sci. USA* **97**, 3370–3375 (2000).
- Gross, J. A. *et al.* *Nature* **404**, 995–999 (2000).
- Kotzin, B. L. *Cell* **85**, 303–306 (1996).
- Morel, L. & Wakeland, E. K. *Curr. Opin. Immunol.* **10**, 718–725 (1998).
- Nagata, S. *Semin. Immunol.* **6**, 3–8 (1994).
- Moore, P. A. *et al.* *Science* **260**, 260–263 (1999).
- Mukhopadhyay, A., Ni, J., Zhai, Y., Yu, G.-H. & Aggarwal, B. B. *J. Biol. Chem.* **274**, 15978–15981 (1999).
- Schneider, P. *et al.* *J. Exp. Med.* **189**, 1747–1756 (1999).
- Von Bülow, G.-U. & Bram, R. J. *Science* **278**, 138–141 (1997).
- Madry, C. *et al.* *Int. Immunol.* **10**, 1693–1702 (1998).
- Chan, O. *et al.* *J. Exp. Med.* **189**, 1639–1647 (1999).

Daedalus

Danger! Men at work

In the modern workplace, masculinity is a grave disadvantage. Few jobs still require the physical strength, competitiveness and courage promoted by testosterone. The modern office rewards submissive teamwork, bland reliability, and careful politeness to superiors, inferiors and customers alike. And a man who even glances speculatively at any of his increasing number of female co-workers risks a charge of sexual harassment or creating a hostile environment. The new business world is fit only for wimps.

So Daedalus is studying the essence of wimpishness. He notes that testosterone can be leaked continuously into the bloodstream from a small skin patch. Remove the patch, and the hormone level declines rapidly. DREADCO biochemists are now inventing an exactly inverse skin patch. One version leaks a testosterone antagonist into the blood; another acts as a trap, sequestering and removing the hormone from the blood passing beneath it. The idea is to reduce the testosterone in a man's blood, if not quite to zero, at any rate to an absolute biological minimum.

The final 'Wimpatch' will effortlessly solve the work-related emotional stresses of so many men. Slapping on a Wimpatch in the morning, the sexiest and most combative man will shrink into the limp, acquiescent role demanded by the modern office or service industry. Peeling it off after work, he will snap back into the bold instinctive style needed for sport, carousing or sexual enterprise.

But the Wimpatch offers its users a still more valuable gain. Male drive and sexuality decline sadly with age, far faster than the slow reduction of blood testosterone levels. Daedalus reckons the testosterone sensors in body and brain become fatigued by decades of steady high concentration, and gradually lose their sensitivity. But a shrewd Wimpatch wearer will fatigue them far less. He will wear his patch regularly at work, and at many other times when he does not actually need to be masculine and sexy. His sensors, used to very low testosterone levels, will not get blunted. So he will retain his male drive and enterprise well into old age. Indeed, when he finally gives up work, and throws his Wimpatch away for the last time, his retirement career will encompass far more sexually vigorous and rewarding pursuits than the usual pathetic pastimes of bowls, golf, gardening and endless TV. **David Jones**

The Further Inventions of Daedalus is published by Oxford University Press.

Focusing hard X-rays with old LPs

A vinyl long-playing record can be used to form a cheap, aberration-free refractive lens.

We have found that two sections cut from a vinyl long-playing record can form a spherical aberration-free refractive lens for hard X-rays. Our manufactured saw-tooth refractive lens has a focal length of 22 cm for 23-keV X-rays. The low cost and short focal length of this lens make it feasible for use in small-scale experiments with conventional X-ray tubes.

Refractive X-ray focusing, although not a new idea¹, has generally been considered inefficient and impracticable^{2,3}. A breakthrough was the compound refractive lens, in which the refractive effect was substantially increased by the combination of many individual cylindrical lenses^{4,5}. From a manufacturing point of view, however, a design with only straight cuts would be preferable.

We have designed such a lens, comprising two identical halves that have a number (N) of regular saw-teeth of height y_s arranged along the optical axis (Fig. 1). At one end the halves touch, and at the opposite end there is a gap ($2y_g$) between them that determines the focal length. Because an X-ray travelling farther away from the axis experiences more surfaces, the total deflection increases with increasing distance from the axis. The lens is tunable and the focal length for a given energy can be varied by a simple mechanical procedure.

In the small-angle approximation, a ray traverses the saw-tooth refractive lens in a straight line parallel to the optical axis. A ray separated by y from the optical axis traverses a thickness of material given by

$$x(y) = y^2 N / (y_g \tan \theta)$$

This is a parabola, with radius of curvature

$$R = y_g \tan \theta / 2N$$

The total length of the lens is $L = 2Ny_s / \tan \theta$, and so $R = (y_g / 2N)(2y_s N / L) = y_g y_s / L$.

Thus we can treat the saw-tooth refractive lens as a single parabolic lens with a focal length of $f = R / \delta$, where δ is the decrement of the real part of the index of refraction from unity (typically 10^{-6} for hard X-rays). The focusing properties are independent of the fixed free parameter θ .

We can apply conventional geometrical optics to calculate the gain of flux. Suppose we wish to focus a beam to a slit with width d_s . The distance between source and lens is s_o , whereas that between lens and slit is s_i . The (de)magnification factor is then given by $M_y = s_i / s_o$. A ray that has lateral displacement y is attenuated by a factor

$$\exp[-x(y)/l] = \exp(-y^2 / 2f\delta l)$$

where l is the X-ray attenuation length. This is a gaussian beam with r.m.s. width

$$\sigma = \sqrt{f\delta l}$$

which is a key lens parameter. It can easily

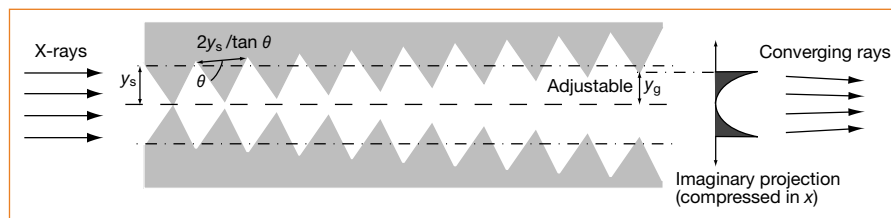


Figure 1 The saw-tooth refractive lens for hard X-rays. Only ten of 300 teeth are shown. The lens bears a striking resemblance to phonograph records, the groove pitch of which is about 180 μm . Measurements of the profile indicate that the depth of the groove is insufficient (about 25 μm) and that there is a large amount of material between the grooves, however, resulting in unnecessary X-ray attenuation. We therefore had a dedicated master cut with a groove depth of 90 μm , from which a vinyl record was pressed. Two 60-mm long sections were cut out to form the lens. With 180 μm separation at the end, the focal length is 218 mm for 23-keV X-rays.

be shown that, to a good approximation, the gain of flux is $G = \sqrt{2\pi}(1 + M_y)\sigma/d_s$.

Because absorption is a limiting factor of the performance of the saw-tooth refractive lens, a material with low atomic number should mitigate it — beryllium and polymethylmethacrylate would be good candidates. We used polyvinyl chloride for focusing as it contains a large fraction of chlorine and provides less gain.

A normal X-ray tube with tungsten anode and focal spot size of 50 μm was used for experimental verification, with a source-to-lens distance of 640 mm and a lens-to-slit distance of 330 mm. The measured gain at 23 keV was 1.7, 15% lower than expected. This was probably due to misalignment and bad surface quality of the cuts. For polymethylmethacrylate, the gain would be three times this (3×1.7).

Because of its simplicity, low cost and easy alignment, the saw-tooth refractive X-ray lens could be valuable as a tool in lab-

oratory-scale arrangements for X-ray applications such as imaging, microscopy, fluorescence and diffraction. As the lens is chromatic, the desired energy can be selected from a broad X-ray spectrum by changing the gap between the lens halves. Two saw-tooth refractive lenses can be put in series to obtain focusing in two directions and squared gain.

Björn Cederström*, **Robert N. Cahn†**, **Mats Danielsson***, **Mats Lundqvist***, **David R. Nygren†**

*Department of Physics, Royal Institute of Technology, SE-104 05 Stockholm, Sweden
e-mail: ceder@particle.kth.se

†Physics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

1. Suehiro, S., Miyaji, H. & Hayashi, H. *Nature* **352**, 385–386 (1991).
2. Michette, A. G. *Nature* **353**, 510 (1991).
3. Yang, B. X. *Nucl. Instrum. Meth. A* **328**, 578–587 (1993).
4. Snigirev, A., Kohn, V., Snigireva, I. & Lengeler, B. *Nature* **384**, 49–51 (1996).
5. Lengeler, B. *et al. J. Appl. Phys.* **84**, 5855–5861 (1998).

Evolution

Symbiotic solution to arsenic contamination

Higher plants that are adapted to living on polluted soils are generally symbiotic with mycorrhizal fungi growing on contaminated sites^{1,2}. It is not known whether these fungi benefit their host plants simply by fulfilling their normal ecological functions³, or by enhancing the plant's resistance to pollutants. Arsenate contamination poses a particular challenge, as this toxin can enter plants through their phosphate transporters⁴, causing mycorrhizal fungi to enhance both phosphate and arsenate uptake in plants⁵. We have found a plant host and its mycorrhizal symbiont that have evolved in parallel to obtain phosphate but exclude arsenate.

We took populations of the ericoid mycorrhizal fungus *Hymenoscyphus ericae* from the roots of *Calluna vulgaris* growing on an arsenic and copper mine spoil and on an uncontaminated heathland site. Arsenate (the arsenic species in soil solution) dose-response curves indicated that plants and fungal populations from the mine were much less sensitive to arsenate than the heathland populations (Fig. 1a). Both partners on the mine spoil have thus evolved resistance to arsenate: the comparatively low sensitivity of the fungus suggests that it may confer added arsenate resistance to the association.

Short-term arsenate- and phosphate-uptake kinetics of the fungus were identical for the mine and heathland populations. Efflux kinetics of arsenic species from mycelia preloaded for 1 h with 0.1 mol m⁻³ arsenate showed that resistant *H. ericae*

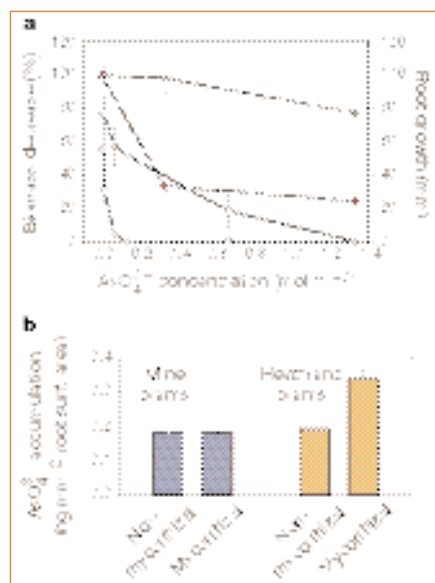


Figure 1 Growth of plant and fungus in the presence of arsenate and the influence of fungal colonization on arsenate accumulation by the host plant. **a**, Biomass of *H. ericae* mycelia after 17 d in medium containing arsenate ($n=24$) and root-lengths of *C. vulgaris* seedlings grown on agar plates containing arsenate for 3 months ($n=9$). Filled symbols, fungus; open symbols, plants. Triangles, mine populations; circles, heathland populations; dotted lines, fungus; solid lines, plants. **b**, Arsenate uptake (24 h) from a 0.1 mmol m⁻³ arsenate solution by 4-month-old sterile *C. vulgaris* seedlings with or without *H. ericae* from site of origin ($n=3$). Graphs show means $\pm$ s.e.

from the mine has enhanced efflux of arsenite (efflux from mine isolates, $14.4 \pm 3.0\%$ h⁻¹; heathland isolates, $6.6 \pm 3.0\%$ h⁻¹ ($P < 0.001$)). *H. ericae* has independently evolved the same strategy as one used by bacteria that reduce arsenate intracellularly to arsenite⁶ for efflux.

Arsenate resistance in *C. vulgaris* is not achieved by exclusion, as we found no alteration to influx or efflux for either arsenate or phosphate. Higher plants normally resist arsenate through downregulation of arsenate/phosphate transporters⁷, but *C. vulgaris* has adopted an alternative mechanism. We suggest that the fungus dominates arsenate/phosphate accumulation, acting as a filter to maintain low plant arsenic levels through arsenite efflux while enhancing plant phosphorus status. When exposed to arsenate for 24 h, heathland *C. vulgaris* inoculated with heathland *H. ericae* accumulated 100% more ($P=0.01$) arsenate than uninoculated heathland plants or mine plants (with and without mine *H. ericae*), which all had comparable arsenate accumulation (Fig. 1b). Infection of mine *C. vulgaris* with mine *H. ericae* did not place the host under additional arsenate stress.

We found that mine and heathland *H. ericae* were constitutively copper resistant, which confers resistance upon non-resistant *C. vulgaris*⁸. In contrast, *H. ericae* and its host must have adapted to arsenate contamination in parallel, by the fungus selectively

accumulating phosphate over arsenate. Our data show that evolution of host and symbiont is fundamental to the colonization of polluted soils by key plant taxa.

J. M. Sharples*†, A. A. Meharg†‡, S. M. Chambers*, J. W. G. Cairney*

*Mycorrhiza Research Group, School of Science, University of Western Sydney, PO Box 10, Kingswood, New South Wales 2747, Australia
e-mail: j.cairney@nepean.uws.edu.au

†Institute of Terrestrial Ecology, Monks Wood, Abbots Ripton, Huntingdon PE17 2LS, UK

‡Present address: Department of Plant and Soil Science, University of Aberdeen, Cruickshank Building, Aberdeen AB24 3UU, UK

1. Macnair, M. R. *New Phytol.* **124**, 541–559 (1993).
2. Cairney, J. W. G. & Meharg, A. A. *Environ. Pollut.* **106**, 169–182 (1999).
3. Smith, S. E. & Read, D. J. *Mycorrhizal Symbiosis* (Academic, San Diego, 1996).
4. Lee, R. B. *Ann. Bot.* **50**, 429–449 (1982).
5. Straker, C. J. & Mitchell, D. T. *New Phytol.* **106**, 129–137 (1987).
6. Silver, S. & Phung, L. T. *Annu. Rev. Microbiol.* **50**, 753–789 (1996).
7. Meharg, A. A. & Macnair, M. R. *J. Exp. Bot.* **43**, 519–524 (1992).
8. Bradley, R., Burt, A. J. & Read, D. J. *Nature* **292**, 335–337 (1981).

Pattern recognition

Tunes and templates

A tune is a succession of musical tones that is easily recognizable when repeated. Here we describe an experimental technique and preliminary results that test a simple theoretical idea about tune recognition.

The theory is that the initial experiences of the melody form a template in the plane defined by pitch and time¹. When the tune is repeated, it is compared with this template, and for the appropriate starting time there will be matches between what is received and what is predicted from the template: the number of these matches is a decision variable determining the detectability of the tune. Detectability can be impaired either by randomly removing some of the notes from the tune, or by adding interfering 'noise' notes to it. The theory predicts how these two types of interference should combine with each other, and we have experimentally tested whether this prediction is correct.

We took a well known melody (the first four bars of *Eine Kleine Nachtmusik* by W. A. Mozart) and mixed random notes with it to find out how many such 'noise' notes could be tolerated while still recognizing the tune. We did the experiment with the whole melody, and also with 1, 2, 4 or 9 notes selected randomly and independently on each trial from the full set of 18, preserving their order and timing. In any one experiment, the number of melody notes was fixed and the task was to identify whether they were present.

It is easiest to understand the experimental design by considering a matrix of 37 rows for the note pitches by 128 columns for their timing. For a given starting time, the melody notes marked 18 elements of the matrix. The masking or noise notes were then scattered randomly within the matrix. The subject's task was to decide whether the stimulus contained part or all of the melody, or alternatively whether it belonged to the set of noise-only notes. For each number of melody notes, Fig. 1 shows the number of masking noise notes that could be tolerated while still allowing 75 per cent correct responses in the single-interval forced-choice trials.

The simplest theory assumes that the starting time of the melody is known. The template acts as a sieve and performance depends on the total number of notes that passed through the holes corresponding to the expected positions of the melody notes. For the 'melody' trials, these numbers exceed those on the 'noise-only' trials by $\Delta\mu$, which is the average number of melody notes in those trials, namely 1, 2, 4, 9 or 18. If σ is the standard deviation of the number that occur in the noise-only trials, then $d' = \Delta\mu/\sigma$ (ref. 2). The statistics of the number of melody-matching notes from the masking noise is binomial; hence σ is equal to $\sqrt{npq}$. In our case, p and q are constant, so σ is proportional to $\sqrt{n}$, the

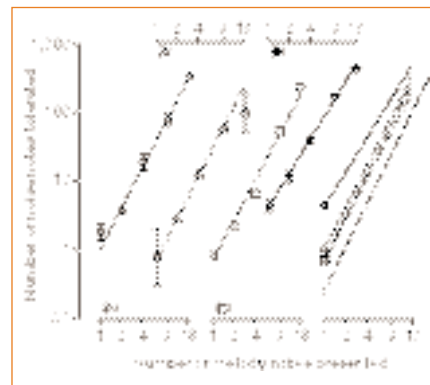


Figure 1 The subjects' task was to distinguish between a 'noise' sequence of notes and a 'melody' sequence. The melody sequence differed from the noise sequence by containing 1, 2, 4, 9 or 18 notes drawn from a melody in addition to noise. The number of noise notes that could be tolerated for 75% correct answers was estimated from 200 trials. The four sets of data points on the left show these results plotted on logarithmic scales; the results for each subject are separated by horizontal shifts. Straight lines were fitted, and these lines with their appropriate symbols are replotted at the right with a single horizontal scale, together with the theoretically predicted line of slope 2 (broken line). Three of the subjects (open symbols) were musically untrained or amateur musicians, and for them the fitted values of the slopes lie reasonably close to the predicted value of 2 (circle, 2.02 ± 0.054 ; triangle, 2.04 ± 0.067 ; square, 1.94 ± 0.12). The fourth subject (filled circles) is a professional musician and was the only subject with absolute pitch. She tolerated more noise notes, her results have less variability and lie on a shallower slope (1.63 ± 0.067) than the others.

square root of the number of noise notes. Thus, for constant d' , n varies as the square of the number of melody notes. On log-log plots we would expect the data values to fall on lines of slope 2.

Figure 1 shows that for three of the subjects this tentative prediction was quite well borne out. The fourth subject was a professional musician: she could tolerate more noise notes, her results were less variable, and they lay on a line of lower slope. This indicates that she had the greatest advantage when only a small number of melody notes were presented, possibly as a result of her well-developed sense of absolute pitch.

Our experiment represents a highly constrained situation because a correct answer can hardly imply recognition of the melody, particularly from only a few melody notes. But it does answer some elementary questions about the basis of musical recognition, and the technique can be developed to test whether it is absolute pitch, the occurrence of particular intervals, or other features of the melody that matter.

These experiments were prompted by results on the detection of biological motion³ and other visual tasks⁴. We believe that they point to another field where the effects of masking noise can help to elucidate the mechanisms of pattern recognition.

Willy Wong*, Horace Barlow†

*Cavendish Laboratory, Cambridge CB3 0HE, UK

†Physiological Laboratory, Cambridge CB2 3EG, UK

e-mail: hbb10@cam.ac.uk

- Warren, R. M. *Auditory Perception*, 126–130 (Cambridge Univ. Press, 1999).
- Swets, J. A., Tanner, W. P. & Birdsall, T. G. in *Signal Detection and Recognition by Human Observers* (ed. Swets, J. A.) 3–57 (Wiley, New York, 1964).
- Neri, P., Morrone, M. C. & Burr, D. C. *Nature* **395**, 894–896 (1998).
- Tripathy, S. P., Mussap, A. J. & Barlow, H. B. *Vision Res.* **39**, 4161–4171 (1999).

Palaeontology

A refugium for relicts?

Luo¹ suggests that the vertebrate fauna from the Yixian Formation (Liaoning Province, China) shows that this region of eastern Asia was a refugium, in which several typically Late Jurassic lineages (compsognathid theropod dinosaurs, 'rhamphorhynchoid' pterosaurs, primitive mammals) survived into the Early Cretaceous¹ (Fig. 1). Data from slightly older sediments in the Japanese Early Cretaceous, however, suggest that the faunal composition of this region can only be partly explained by the concept of a refugium.

The Kuwajima Formation of Ishikawa Prefecture, central Japan, is yielding an important Early Cretaceous vertebrate fauna. This unit is a lateral equivalent of the



Figure 1 Stratigraphic ranges of clades that include taxa recovered from the Yixian Formation, China, and the Kuwajima and Itsuki Formations, Japan^{1,4,9}. Data on *Camptosaurus* and *Echinodon* are from ref. 13. Arrows, lineage extends beyond the time range shown here; solid bars, first and last occurrences. Al, Albian; Ap, Aptian; Ba, Bathonian; Be, Berriasian; Br, Barremian; Ca, Callovian; Ha, Hauterivian; Ki, Kimmeridgian; Ox, Oxfordian; Ti, Tithonian; Va, Valanginian.

Okurodani Formation that outcrops in neighbouring Gifu Prefecture². Stratigraphic, biostratigraphic and radiometric data show that the Okurodani Formation is basal Cretaceous (Valanginian or Hauterivian) in age³. The Kuwajima Formation has yielded more than one hundred isolated teeth of a new genus of tritylodontid synapsid⁴. Before these discoveries, tritylodontids were thought to have become extinct sometime in the Middle or early Late Jurassic, as the youngest-known tritylodontid (*Bienotheroides*) was recovered from late Middle Jurassic deposits. This discovery supports the concept of an East Asian refugium, but other evidence suggests that different factors may have had an equally strong influence on faunal composition.

A theropod dinosaur referable to the unnamed clade Oviraptorosauria + Therizinosauroidea⁵ has also been found in the tritylodontid locality. This clade is best known from the Late Cretaceous of mainland Asia, although several taxa referable to this clade are known from the late Early Cretaceous of Liaoning (*Beipiaosaurus*⁶ and *Caudipteryx*⁷), and possibly from the Early Jurassic of Yunnan Province, China⁸. The Japanese material, consisting of a single manual ungual (Fig. 2) with a pronounced posterodorsal lip (a feature synapomorphic of this group of theropods⁵), is one of the earliest representatives of this group. The Itsuki Formation of Fukui Prefecture, a lateral equivalent of the Okurodani and Kuwajima Formations², has produced an isolated tyrannosaurid tooth, identifiable by its D-shaped cross-section — a synapomorphy of tyrannosaurids⁹.

These Japanese discoveries, combined with the presence of late Early Cretaceous taxa in the Yixian Formation (such as the ornithischian dinosaur *Psittacosaurus*¹⁰), suggest that several dinosaur clades (such as tyrannosaurids and psittacosaurids) may have originated and diversified in eastern



Figure 2 Manual ungual of a theropod dinosaur from the Kuwajima Formation (Valanginian or Hauterivian) of Shiramine, Ishikawa Prefecture, Japan. Note the prominent lip posterodorsal to the articular surface of the ungual, a synapomorphy of the clade Oviraptorosauria + Therizinosauroidea⁵. Scale bar, 5 mm.

Asia while a number of other lineages (tritylodontid synapsids, compsognathid dinosaurs and 'rhamphorhynchoid' pterosaurs) persisted in this region. Moreover, the presence of hypsilophodontid and iguanodontid ornithopod dinosaurs in the Japanese Early Cretaceous¹¹ suggests faunal connections with western Asia and Europe. The historical biogeography of this region appears to be much more complex than was thought previously.

Alternatively, the so-called relict taxa in eastern Asia may indicate that faunal turnover at the Jurassic–Cretaceous boundary was not as marked as has been suggested¹². The presence of camptosaurid (*Camptosaurus*) and heterodontosaurid (*Echinodon*) ornithopods in European Early Cretaceous faunas¹³ indicates faunal similarities to the Late Jurassic Morrison Formation of North America. The presence of 'Late Jurassic' taxa in eastern Asia may simply represent another example of this more gradual Jurassic–Cretaceous faunal transition (Fig. 1), although more evidence is needed to distinguish between these alternatives.

Makoto Manabe*, Paul M. Barrett†, Shinji Isaji‡

*Department of Geology, National Science Museum, 3-23-1 Hyakunin-cho, Shinjuku-ku, Tokyo 169-0073, Japan

†Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

‡Chiba Prefectural Museum of Natural History, Chiba 260-0682, Japan

- Luo, Z. *Nature* **400**, 23–25 (1999).
- Maeda, S. J. *Collect. Arts Sci. Chiba Univ.* **3**, 369–426 (1961).
- Evans, S. E. et al. *New Mexico Mus. Nat. Hist. Sci. Bull.* **14**, 183–186 (1998).
- Setoguchi, T., Matsuoka, H. & Matsuda, M. in *Proc. 7th Annu. Meet. Chinese Soc. Vert. Paleontol.* (eds Wang, Y.-Q. & Deng, T.) 117–124 (China Ocean, Beijing, 1999).
- Makovicky, P. J. & Sues, H.-D. *Am. Mus. Novitates* **3240**, 1–27 (1998).
- Xu, X., Tang, Z.-L. & Wang, X.-L. *Nature* **399**, 350–354 (1999).
- Sereno, P. C. *Science* **284**, 2137–2147 (1999).
- Zhao, X.-J. & Xu, X. *Nature* **394**, 234–235 (1998).
- Manabe, M. J. *Paleontol.* **73**, 1176–1178 (1999).
- Xu, X. & Wang, X.-L. *Vertebrata Palasiatica* **36**, 147–158 (1998).
- Manabe, M. & Hasegawa, Y. in *6th Symp. Mesozoic Terrest. Ecosyst. Biotas* (eds Sun, A.-L. & Wang, Y.-Q.) 179 (China Ocean, 1995).
- Bakker, R. T. *Nature* **274**, 661–663 (1978).
- Norman, D. B. & Barrett, P. M. *Spec. Pap. Palaeontol.* (in the press).

A flat Universe from high-resolution maps of the cosmic microwave background radiation

P. de Bernardis¹, P. A. R. Ade², J. J. Bock³, J. R. Bond⁴, J. Borrill^{5,12}, A. Boscaleri⁶, K. Coble⁷, B. P. Crill⁸, G. De Gasperis⁹, P. C. Farese⁷, P. G. Ferreira¹⁰, K. Ganga^{8,11}, M. Giacometti¹, E. Hivon⁸, V. V. Hristov³, A. Iacoangeli¹, A. H. Jaffe¹², A. E. Lange⁸, L. Martinis¹³, S. Masi¹, P. V. Mason⁸, P. D. Mauskopf^{14,15}, A. Melchiorri¹, L. Miglio¹⁶, T. Montroy⁷, C. B. Netterfield¹⁶, E. Pascale⁶, F. Piacentini¹, D. Pogosyan⁴, S. Prunet⁴, S. Rao¹⁷, G. Romeo¹⁷, J. E. Ruhl⁷, F. Scaramuzzi¹³, D. Sforna¹ & N. Vittorio⁹

The blackbody radiation left over from the Big Bang has been transformed by the expansion of the Universe into the nearly isotropic 2.73 K cosmic microwave background. Tiny inhomogeneities in the early Universe left their imprint on the microwave background in the form of small anisotropies in its temperature. These anisotropies contain information about basic cosmological parameters, particularly the total energy density and curvature of the Universe. Here we report the first images of resolved structure in the microwave background anisotropies over a significant part of the sky. Maps at four frequencies clearly distinguish the microwave background from foreground emission. We compute the angular power spectrum of the microwave background, and find a peak at Legendre multipole $l_{\text{peak}} = (197 \pm 6)$, with an amplitude $\Delta T_{200} = (69 \pm 8) \mu\text{K}$. This is consistent with that expected for cold dark matter models in a flat (euclidean) Universe, as favoured by standard inflationary models.

Photons in the early Universe were tightly coupled to ionized matter through Thomson scattering. This coupling ceased about 300,000 years after the Big Bang, when the Universe cooled sufficiently to form neutral hydrogen. Since then, the primordial photons have travelled freely through the Universe, redshifting to microwave frequencies as the Universe expanded. We observe those photons today as the cosmic microwave background (CMB). An image of the early Universe remains imprinted in the temperature anisotropy of the CMB. Anisotropies on angular scales larger than $\sim 2^\circ$ are dominated by the gravitational redshift the photons undergo as they leave the density fluctuations present at decoupling^{1,2}. Anisotropies on smaller angular scales are enhanced by oscillations of the photon–baryon fluid before decoupling³. These oscillations are driven by the primordial density fluctuations, and their nature depends on the matter content of the Universe.

In a spherical harmonic expansion of the CMB temperature field, the angular power spectrum specifies the contributions to the fluctuations on the sky coming from different multipoles l , each corresponding to the angular scale $\theta = \pi/l$. Density fluctuations over spatial scales comparable to the acoustic horizon at decoupling produce a peak in the angular power spectrum of the CMB, occurring at multipole l_{peak} . The exact value of l_{peak} depends on both the linear size of the acoustic horizon and on the angular diameter distance from the observer to decoupling. Both these quantities are sensitive to a number of cosmological parameters (see, for example, ref. 4), but l_{peak} primarily depends on the total density of the Universe, Ω_0 . In models with a density Ω_0 near 1,

$l_{\text{peak}} \approx 200/\Omega_0^{1/2}$. A precise measurement of l_{peak} can efficiently constrain the density and thus the curvature of the Universe.

Observations of CMB anisotropies require extremely sensitive and stable instruments. The DMR⁵ instrument on the COBE satellite mapped the sky with an angular resolution of $\sim 7^\circ$, yielding measurements of the angular power spectrum at multipoles $l < 20$. Since then, experiments with finer angular resolution^{6–16} have detected CMB fluctuations on smaller scales and have produced evidence for the presence of a peak in the angular power spectrum at $l_{\text{peak}} \approx 200$.

Here we present high-resolution, high signal-to-noise maps of the CMB over a significant fraction of the sky, and derive the angular power spectrum of the CMB from $l = 50$ to 600. This power spectrum is dominated by a peak at multipole $l_{\text{peak}} = (197 \pm 6)$ (1σ error). The existence of this peak strongly supports inflationary models for the early Universe, and is consistent with a flat, euclidean Universe.

The instrument

The Boomerang (balloon observations of millimetric extragalactic radiation and geomagnetics) experiment is a microwave telescope that is carried to an altitude of ~ 38 km by a balloon. Boomerang combines the high sensitivity and broad frequency coverage pioneered by an earlier generation of balloon-borne experiments with the long (~ 10 days) integration time available in a long-duration balloon flight over Antarctica. The data described here were obtained with a focal plane array of 16 bolometric detectors cooled to 0.3 K. Single-mode feedhorns provide two $18'$ full-width at half-maximum (FWHM) beams at 90 GHz and two $10'$ FWHM beams at 150 GHz. Four multi-band photometers each provide a $10.5'$, $14'$ and $13'$ FWHM beam at 150, 240 and 400 GHz respectively. The average in-flight sensitivity to CMB anisotropies was 140, 170, 210 and $2,700 \mu\text{K s}^{1/2}$ at 90, 150, 240 and 400 GHz, respectively. The entire optical system is heavily baffled against terrestrial radiation. Large sunshields improve rejection of radiation from $>60^\circ$ in azimuth from the telescope boresight. The rejection has been measured to be greater than 80 dB at all angles occupied by the Sun during the CMB observations. Further details on the instrument can be found in refs 17–21.

¹ Dipartimento di Fisica, Università di Roma “La Sapienza”, P.le A. Moro 2, 00185 Roma, Italy.

² Department of Physics, Queen Mary and Westfield College, Mile End Road, London E1 4NS, UK.

³ Jet Propulsion Laboratory, Pasadena, California 91109, USA. ⁴ CITA University of Toronto, Toronto M5S 3H8, Canada. ⁵ NERSC-LBNL, Berkeley, California 94720, USA. ⁶ IROE-CNR, Via Panciatichi 64, 50127 Firenze, Italy. ⁷ Department of Physics, University of California at Santa Barbara, Santa Barbara, California 93106, USA. ⁸ California Institute of Technology, Mail Code 59-33, Pasadena, California 91125, USA. ⁹ Dipartimento di Fisica, Università di Roma Tor Vergata, Via della Ricerca Scientifica 1, 00133 Roma, Italy. ¹⁰ Astrophysics, University of Oxford, Keble Road, OX1 3RH, UK. ¹¹ PCC, Collège de France, 11 pl. Marcelin Berthelot, 75231 Paris Cedex 05, France. ¹² Center for Particle Astrophysics, University of California at Berkeley, 301 Le Conte Hall, Berkeley, California 94720, USA. ¹³ ENEA Centro Ricerche di Frascati, Via E. Fermi 45, 00044 Frascati, Italy. ¹⁴ Physics and Astronomy Department, Cardiff University, Cardiff CF2 3YB, UK. ¹⁵ Department of Physics and Astronomy, University of Massachusetts, Amherst, Massachusetts 01003, USA. ¹⁶ Department of Physics and Astronomy, University of Toronto, Toronto M5S 3H8, Canada. ¹⁷ Istituto Nazionale di Geofisica, Via di Vigna Murata 605, 00143, Roma, Italy.

Observations

Boomerang was launched from McMurdo Station (Antarctica) on 29 December 1998, at 3:30 GMT. Observations began 3 hours later, and continued uninterrupted during the 259-hour flight. The payload approximately followed the 79°S parallel at an altitude that varied daily between 37 and 38.5 km, returning within 50 km of the launch site.

We concentrated our observations on a target region, centred at roughly right ascension (RA) 5h, declination (dec.) -45° , that is uniquely free of contamination by thermal emission from inter-

stellar dust²² and that is approximately opposite the Sun during the austral summer. We mapped this region by repeatedly scanning the telescope through 60° at fixed elevation and at constant speed. Two scan speeds (1° s^{-1} and 2° s^{-1} in azimuth) were used to facilitate tests for systematic effects. As the telescope scanned, degree-scale variations in the CMB generated sub-audio frequency signals in the output of the detector²³. The stability of the detector system was sufficient to allow sensitive measurements on angular scales up to tens of degrees on the sky. The scan speed was sufficiently rapid with respect to sky rotation that identical structures were observed by

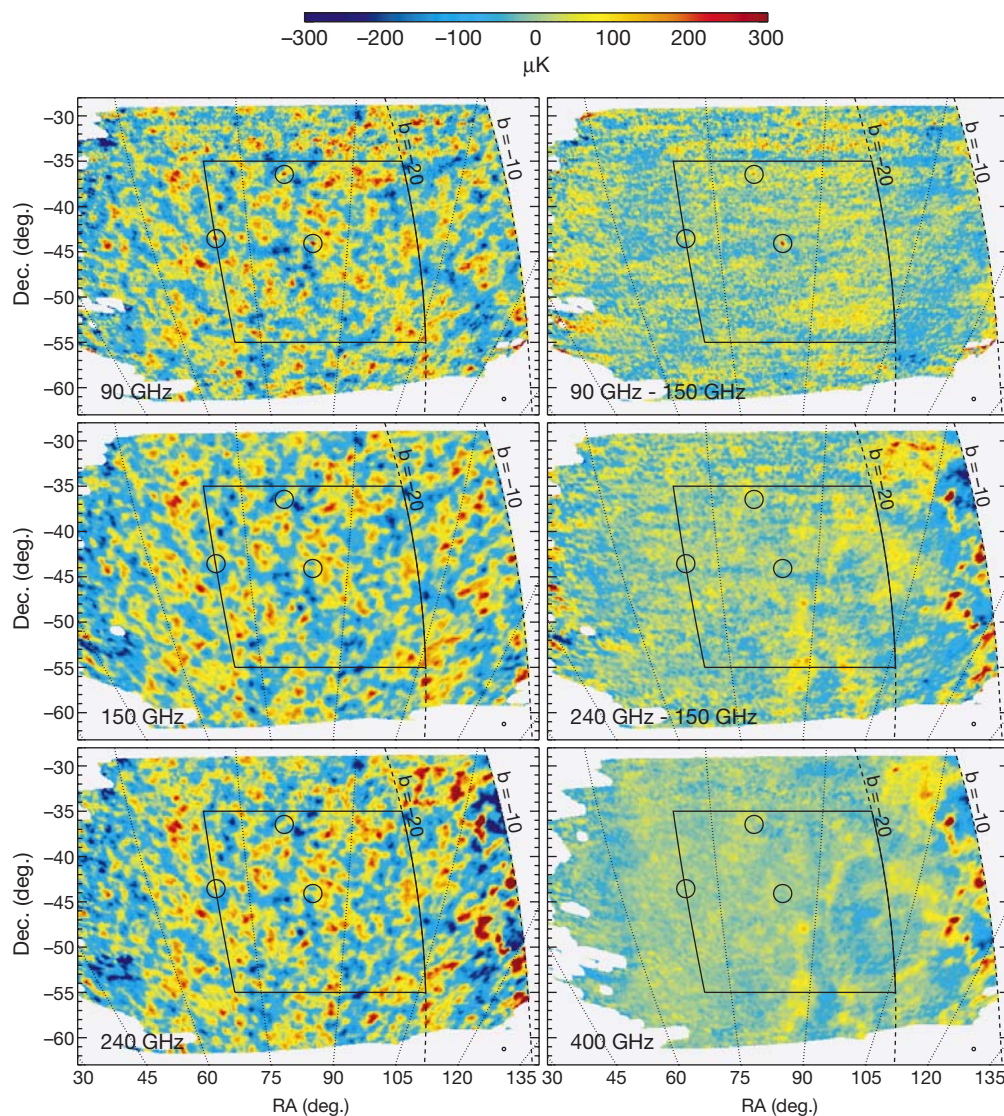


Figure 1 Boomerang sky maps (equatorial coordinates). The sky maps at 90, 150 and 240 GHz (left panels) are shown with a common colour scale, using a thermodynamic temperature scale chosen such that CMB anisotropies will have the same amplitude in the three maps. Only the colour scale of the 400 GHz map (bottom right) is 14 times larger than the others: this has been done to facilitate comparison of emission from interstellar dust (ISD), which dominates this map, with ISD emission present in the lower-frequency maps. The maps at 90 and 400 GHz are each from a single detector, while maps at 150 and 240 GHz have each been obtained by co-adding data from three detectors. For purposes of presentation, the maps have been smoothed with gaussian filters to obtain FWHM effective resolution of $22.5'$ (small circle in the bottom right side of each panel). Structures along the scan direction larger than 10° are not present in the maps. Several features are immediately evident. Most strikingly, the maps at 90, 150 and 240 GHz are dominated by degree-scale structures that fill the map, have well-correlated morphology and are identical in amplitude in all three maps. These structures are not visible at 400 GHz. The 400 GHz map is dominated by diffuse emission which is correlated with the

ISD emission mapped by IRAS/DIRBE²². This emission is strongly concentrated towards the right-hand edge of the maps, near the plane of the Galaxy. The same structures are evident in the 90, 150 and 240 GHz maps at Galactic latitude $b > -15^\circ$, albeit with an amplitude that decreases steeply with decreasing frequency. The large-scale gradient evident especially near the right edge of the 240 GHz map is a result of high-pass-filtering the very large signals near the Galactic plane (not shown). This effect is negligible in the rest of the map. The two top right panels show maps constructed by differencing the 150 and 90 GHz maps and the 240 and 150 GHz maps. The difference maps contain none of the structures that dominate the maps at 90, 150 and 240 GHz, indicating that these structures do indeed have the ratios of brightness that are unique to the CMB. The morphology of the residual structures in the 240–150 GHz map is well-correlated with the 400 GHz map, as is expected if the residuals are due to the ISD emission. Three compact sources of emission are visible in the lower-frequency maps, as indicated by the circles. These are known radio-bright quasars from the SEST pointing catalogue at 230 GHz. The boxed area has been used for computing the angular power spectrum shown in Fig. 2.

detectors in the same row in each scan. Detectors in different rows observed the same structures delayed in time by a few minutes.

At intervals of several hours, the telescope elevation was interchanged between 40°, 45° and 50° in order to increase the sky coverage and to provide further systematic tests. Sky rotation caused the scan centre to move and the scan direction to rotate on the celestial sphere. A map from a single day at a single elevation covered roughly 22° in declination and contained scans rotated by $\pm 11^\circ$ on the sky, providing a cross-linked scan pattern. Over most of the region mapped, each sky pixel was observed many times on different days, both at 1° s^{-1} and 2° s^{-1} scan speed, with different topography, solar elongation and atmospheric conditions, allowing strong tests for any contaminating signal not fixed on the celestial sphere.

The pointing of the telescope has been reconstructed with an accuracy of $2'$ r.m.s. using data from a Sun sensor and rate gyros. This precision has been confirmed by analysing the observed positions of bright compact H II regions in the Galactic plane (RCW38²⁴, RCW57, IRAS08576 and IRAS1022) and of radio-bright point sources visible in the target region (the QSO 0483–436, the BL Lac object 0521–365 and the blazar 0537–441).

Calibrations

The beam pattern for each detector was mapped before flight using a thermal source. The main lobe at 90, 150 and 400 GHz is accurately modelled by a gaussian function. The 240 GHz beams are well modelled by a combination of two gaussians. The beams have small shoulders (less than 1% of the total solid angle), due to aberrations in the optical system. The beam-widths were confirmed in flight via observations of compact sources. By fitting radial profiles to these sources we determine the effective angular resolution, which includes the physical beamwidth and the effects of the $2'$ r.m.s. pointing jitter. The effective FWHM angular resolution of the 150 GHz data that we use here to calculate the CMB power spectrum is $(10 \pm 1)'$, where the error is dominated by uncertainty in the pointing jitter.

We calibrated the 90, 150 and 240 GHz channels from their measured response to the CMB dipole. The dipole anisotropy has been accurately (0.7%) measured by COBE-DMR²⁵, fills the beam and has the same spectrum as the CMB anisotropies at smaller angular scales, making it the ideal calibrator for CMB experiments. The dipole signal is typically ~ 3 mK peak-to-peak in each 60° scan, much larger than the detector noise, and appears in the output of the detectors at $f = 0.008$ Hz and $f = 0.016$ Hz in the 1° s^{-1} and 2° s^{-1} scan speeds, respectively. The accuracy of the calibration is dominated by two systematic effects: uncertainties in the low-frequency transfer function of the electronics, and low-frequency, scan-synchronous signals. Each of these is significantly different at the two scan speeds. We found that the dipole-fitted amplitudes

derived from separate analysis of the 1° s^{-1} and 2° s^{-1} data agree to within $\pm 10\%$ for every channel, and thus we assign a 10% uncertainty in the absolute calibration.

From detector signals to CMB maps

The time-ordered data comprises 5.4×10^7 16-bit samples for each channel. These data are flagged for cosmic-ray events, elevation changes, focal-plane temperature instabilities, and electromagnetic interference events. In general, about 5% of the data for each channel are flagged and not used in the subsequent analysis. The gaps resulting from this editing are filled with a constrained realization of noise in order to minimize their effect in the subsequent filtering of the data. The data are deconvolved by the bolometer and electronics transfer functions to recover uniform gain at all frequencies.

The noise power spectrum of the data and the maximum-likelihood maps^{26–28} were calculated using an iterative technique²⁹ that separates the sky signal from the noise in the time-ordered data. In this process, the statistical weights of frequencies corresponding to angular scales larger than 10° on the sky are set to zero to filter out the largest-scale modes of the map. The maps were pixelized according to the HEALPix pixelization scheme³⁰.

Figure 1 shows the maps obtained in this way at each of the four frequencies. The 400 GHz map is dominated by emission from interstellar dust that is well correlated with that observed by the IRAS and COBE/DIRBE satellites. The 90, 150 and 240 GHz maps are dominated by degree-scale structures that are resolved with high signal-to-noise ratio. A qualitative but powerful test of the hypothesis that these structures are CMB anisotropy is provided by subtracting one map from another. The structures evident in all

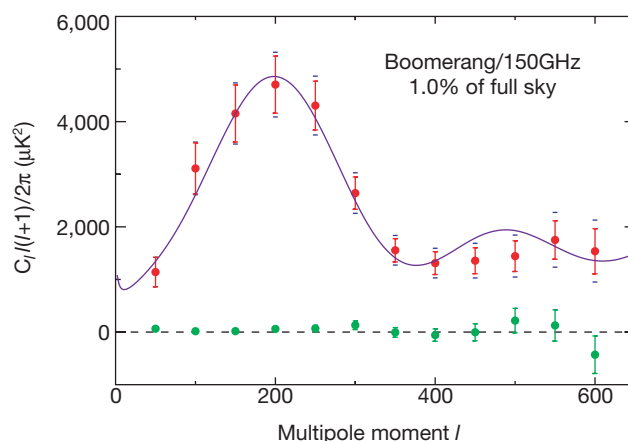


Figure 2 Angular power spectrum measured by Boomerang at 150 GHz. Each point is the power averaged over $\Delta l = 50$ and has negligible correlations with the adjacent points. The error bars indicate the uncertainty due to noise and cosmic/sampling variance. The errors are dominated by cosmic/sampling variance at $l < 350$; they grow at large l due to the signal attenuation caused by the combined effects³⁹ of the $10'$ beam and the $14'$ pixelization (0.87 at $l = 200$ and 0.33 at $l = 600$). The current $\pm 10\%$ uncertainty in the calibration corresponds to an overall re-scaling of the y-axis by $\pm 20\%$, and is not shown. The current $1'$ uncertainty in the angular resolution of the measurement creates an additional uncertainty—indicated by the distance between the ends of the red error bars and the blue horizontal lines—that is completely correlated and is largest (11%) at $l = 600$. The green points show the power spectrum of a difference map obtained as follows. We divided the data into two parts corresponding to the first and second halves of the timestream. We made two maps (A and B) from these halves, and the green points show the power spectrum computed from the difference map, $(A-B)/2$. Signals originating from the sky should disappear in this map, so this is a test for contamination in the data (see text). The solid curve has parameters $(\Omega_b, \Omega_m, \Omega_\Lambda, n_s, h) = (0.05, 0.31, 0.75, 0.95, 0.70)$. It is the best-fit model for the Boomerang test flight data^{15,16}, and is shown for comparison only. The model that best fits the new data reported here will be presented elsewhere.

Table 1 Angular power spectrum of CMB anisotropy

l range	150 GHz ((1st half) + [2nd half])/2	150 GHz ((1st half) – [2nd half])/2
26–75	1,140 $\pm$ 280	63 $\pm$ 32
76–125	3,110 $\pm$ 490	16 $\pm$ 20
126–175	4,160 $\pm$ 540	17 $\pm$ 28
176–225	4,700 $\pm$ 540	59 $\pm$ 44
226–275	4,300 $\pm$ 460	68 $\pm$ 59
276–325	2,640 $\pm$ 310	130 $\pm$ 82
326–375	1,550 $\pm$ 220	–7 $\pm$ 92
376–425	1,310 $\pm$ 220	–60 $\pm$ 120
426–475	1,360 $\pm$ 250	0 $\pm$ 160
476–525	1,440 $\pm$ 290	220 $\pm$ 230
526–575	1,750 $\pm$ 370	130 $\pm$ 300
576–625	1,540 $\pm$ 430	–430 $\pm$ 360

Shown are measurements of the angular power spectrum of the cosmic microwave background at 150 GHz, and tests for systematic effects. The values listed are for $\Delta T^2 = l(l+1) \text{ c}_l^2 / 2\pi$, in μK^2 . Here $c_l = \langle \theta_l^2 \rangle$, and a_{lm} are the coefficients of the spherical harmonic decomposition of the CMB temperature field: $\Delta T(\theta, \phi) = \sum_{lm} a_{lm} Y_{lm}(\theta, \phi)$. The stated 1σ errors include statistical and cosmic/sample variance, and do not include a 10% calibration uncertainty.

three maps disappear in both the 90–150 GHz difference and in the 240–150 GHz difference, as expected for emission that has the same spectrum as the CMB dipole anisotropy used to calibrate the maps.

To quantify this conclusion, we performed a ‘colour index’ analysis of our data. We selected the $\sim 18,000$ 14’ pixels at Galactic latitude $b < -15^\circ$, and made scatter plots of 90 GHz versus 150 GHz and 240 GHz versus 150 GHz. A linear fit to these scatter plots gives slopes of 1.00 ± 0.15 and 1.10 ± 0.16 , respectively (including our present 10% calibration error), consistent with a CMB spectrum. For comparison, free–free emission with spectral index -2.35 would produce slopes of 2.3 and 0.85, and was therefore rejected with $>99\%$ confidence; emission from interstellar dust with temperature $T_d = 15$ K and spectral index of emissivity $\alpha = 1$ would produce slopes of 0.40 and 2.9. For any combination of $T_d > 7$ K and $1 < \alpha < 2$, the dust hypothesis is rejected with $>99\%$ confidence. We conclude that the dominant source of structure that we detect at 90, 150 and 240 GHz is CMB anisotropy.

We further argue that the 150 GHz map at $b < -15^\circ$ is free of significant contamination by any known astrophysical foreground. Galactic synchrotron and free–free emission is negligible at this frequency³¹. Contamination from extragalactic point sources is also small³²; extrapolation of fluxes from the PMN survey³³ limits the contribution by point sources (including the three above-mentioned radio-bright sources) to the angular power spectrum derived below to $<0.7\%$ at $l = 200$ and $<20\%$ at $l = 600$. The astrophysical foreground that is expected to dominate at 150 GHz is thermal emission from interstellar dust. We placed a quantitative limit on this source of contamination as follows. We assumed that dust properties are similar at high ($b < -20^\circ$) and moderate ($-20^\circ < b < -5^\circ$) Galactic latitudes. We selected the pixels at moderate Galactic latitudes and correlated the structure observed in each of our four bands with the IRAS/DIRBE map, which is dominated by dust in cirrus clouds. The best-fit slope of each of the scatter plots measures the ratios of the dust signal in the Boomerang channels to the dust signal in the IRAS/DIRBE map.

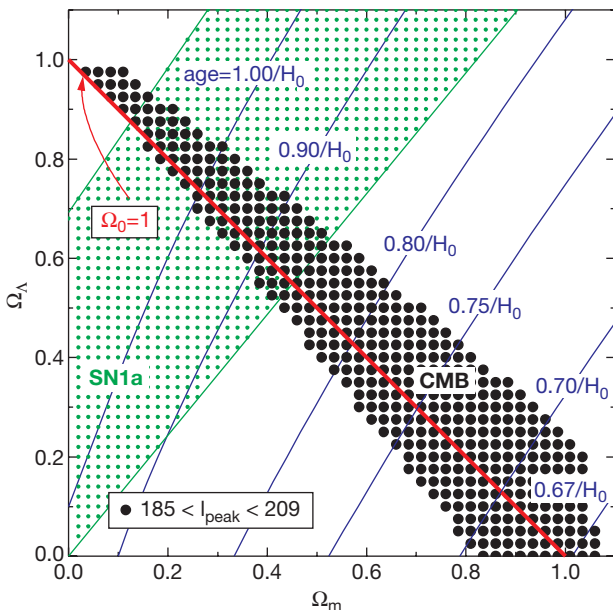


Figure 3 Observational constraints on Ω_m and Ω_Λ . All the cosmological models (from our data base) consistent with the position of the peak in the angular power spectrum measured by Boomerang (95% confidence intervals) define an ‘allowed’ region in the $\Omega_m - \Omega_\Lambda$ plane (marked by large black dots). Such a region is elongated around the $\Omega_0 = 1$ line identifying a flat geometry, euclidean Universe. The blue lines define the age of the Universe for the considered models. The green-dotted region is consistent (95% confidence contour) with the recent results of the high-redshift supernovae surveys^{40,41}.

We found that the 400 GHz map is very well correlated to the IRAS/DIRBE map, and that dust at $b < -20^\circ$ can account for at most 10% of the signal variance at 240 GHz, 3% at 150 GHz and 0.5% at 90 GHz.

Angular power spectra

We compared the angular power spectrum of the structures evident in Fig. 1 with theoretical predictions. In doing so, we separated and removed the power due to statistical noise and systematic artefacts from the power due to the CMB anisotropies in the maps. The maximum-likelihood angular power spectrum of the maps was computed using the MADCAP³⁴ software package, whose algorithms fully take into account receiver noise and filtering.

Full analysis of our entire data set is under way. Because of the computational intensity of this process, we report here the results of a complete analysis of a limited portion of the data chosen as follows. We analysed the most sensitive of the 150 GHz detectors. We restricted the sky coverage to an area with $RA > 70^\circ$, $b < -20^\circ$ and $-55^\circ < \text{dec.} < -35^\circ$, and we used only the $\sim 50\%$ of the data from this detector that was obtained at a scan speed of 1° s^{-1} . We used a relatively coarse pixelization of 8,000 14-arcmin pixels as a compromise between computation speed and coverage of high multipoles. Finally, we limited our analysis to $l \leq 600$ for which the effects of pixel shape and size and our present uncertainty in the beam size ($1'$) are small and can be accurately modelled.

The angular power spectrum determined in this way is shown in Fig. 2 and reported in Table 1. The power spectrum is dominated by a peak at $l_{\text{peak}} \approx 200$, as predicted by inflationary cold dark matter models. These models additionally predict the presence of secondary peaks. The data at high l limit the amplitude, but do not exclude the presence, of a secondary peak. The errors in the angular power spectrum are dominated at low multipoles ($l \lesssim 350$) by the cosmic/sampling variance, and at higher multipoles by detector noise.

The CMB angular power spectrum shown in Fig. 2 was derived from 4.1 days of observation. As a test of the stability of the result, we made independent maps from the first and second halves of these data. The payload travels several hundred kilometres, and the Sun moves 2° on the sky, between these maps. Comparing them provides a stringent test for contamination from sidelobe pickup and thermal effects. The angular power spectrum calculated for the difference map is shown in Fig. 2. The reduced χ^2 of this power spectrum with respect to zero signal is 1.11 (12 degrees of freedom), indicating that the difference map is consistent with zero contamination.

A peak at $l \approx 200$ implies a flat Universe

The location of the first peak in the angular power spectrum of the CMB is well measured by the Boomerang data set. From a parabolic fit to the data at $l = 50$ to 300 in the angular power spectrum, we find $l_{\text{peak}} = (197 \pm 6)$ (1σ error). The parabolic fit does not bias the determination of the peak multipole: applying this method to Monte Carlo realizations of theoretical power spectra we recover the correct peak location for a variety of cosmological models. Finally, the peak location is independent of the details of the data calibration, which obviously affect only the height of the peak and not its location. The height of the peak is $\Delta T_{200} = (69 \pm 4) \pm 7 \mu\text{K}$ (1σ statistical and calibration errors, respectively).

The data are inconsistent with current models based on topological defects (see, for example, ref. 35) but are consistent with a subset of cold dark matter models. We generated a database of cold dark matter models^{36,37}, varying six cosmological parameters (the range of variation is given in parentheses): the non-relativistic matter density, Ω_m (0.05–2); the cosmological constant, Ω_Λ (0–1); the Hubble constant, h (0.5–0.8); the baryon density, $h^2\Omega_b$ (0.013–0.025), the primordial scalar spectral index, n_s (0.8–1.3); and the overall normalization A (free parameter) of the primordial density fluctuation power spectrum. We compared these models with the power spectrum we report here to place constraints on

allowed regions in this 6-parameter space. In Figure 3 we mark with large black dots the region of the Ω_m – Ω_Λ plane where some combination of the remaining four parameters within the ranges defined by our model space gives a power spectrum consistent with our 95% confidence interval for l_{peak} . This region is quite narrow, and elongated along the ‘flat Universe’ line $\Omega_m + \Omega_\Lambda = 1$. The width of this region is determined by degeneracy in the models, which produce closely similar spectra for different values of the parameters³⁸. We further evaluated the likelihood of the models given the Boomerang measurement and the same priors (constraints on the values of the cosmological parameters) as in ref. 16. Marginalizing over all the other parameters, we found the following 95% confidence interval for $\Omega_0 = \Omega_m + \Omega_\Lambda$: $0.88 < \Omega_0 < 1.12$. This provides evidence for a euclidean geometry of the Universe. Our data clearly show the presence of power beyond the peak at $l = 197$, corresponding to smaller-scale structures. The consequences of this fact will be fully analysed elsewhere. □

Received 24 March; accepted 3 April 2000.

1. Sachs, R. K. & Wolfe, A. M. Perturbations of a cosmological model and angular variations of the microwave background. *Astrophys. J.* **147**, 73–90 (1967).
2. Weinberg S., *Gravitation and Cosmology* (Wiley & Sons, New York, 1972).
3. Hu, W., Sugiyama, N. & Silk, J. The physics of cosmic microwave background anisotropies. *Nature* **386**, 37–43 (1997).
4. Bond, J. R., Efstathiou, G. & Tegmark, M. Forecasting cosmic parameter errors from microwave background anisotropy experiments. *Mon. Not. R. Astron. Soc.* **291**, L33–L41 (1997).
5. Hinshaw, G. *et al.* Band power spectra in the COBE-DMR four-year anisotropy map. *Astrophys. J.* **464**, L17–L20 (1996).
6. Scott, P. F. *et al.* Measurement of structure in the cosmic background radiation with the Cambridge cosmic anisotropy telescope. *Astrophys. J.* **461**, L1–L4 (1996).
7. Netterfield, C. B. *et al.* A measurement of the angular power spectrum of the anisotropy in the cosmic microwave background. *Astrophys. J.* **474**, 47–66 (1997).
8. Leitch, E. M. *et al.* A measurement of anisotropy in the cosmic microwave background on 7–22 arcminute scales. *Astrophys. J.* (submitted); also as preprint astro-ph/9807312 at (<http://xxx.lanl.gov>) (1998).
9. Wilson, G. W. *et al.* New CMB power spectrum constraints from MSAMI. *Astrophys. J.* (submitted); also as preprint astro-ph/9902047 at (<http://xxx.lanl.gov>) (1999).
10. Baker, J. C. *et al.* Detection of cosmic microwave background structure in a second field with the cosmic anisotropy telescope. *Mon. Not. R. Astron. Soc.* (submitted); also as preprint astro-ph/9904415 at (<http://xxx.lanl.gov>) (1999).
11. Peterson, J. B. *et al.* First results from Viper: detection of small-scale anisotropy at 40 GHz. Preprint astro-ph/9910503 at (<http://xxx.lanl.gov>) (1999).
12. Coble, K. *et al.* Anisotropy in the cosmic microwave background at degree angular scales: Python V results. *Astrophys. J.* **519**, L5–L8 (1999).
13. Torbet, E. *et al.* A measurement of the angular power spectrum of the microwave background made from the high Chilean Andes. *Astrophys. J.* **521**, 79–82 (1999).
14. Miller, A. D. *et al.* A measurement of the angular power spectrum of the CMB from $l = 100$ to 400. *Astrophys. J.* (submitted); also as preprint astro-ph/9906421 at (<http://xxx.lanl.gov>) (1999).
15. Mauskopf, P. *et al.* Measurement of a peak in the CMB power spectrum from the test flight of BOOMERanG. *Astrophys. J.* (submitted); also as preprint astro-ph/9911444 at (<http://xxx.lanl.gov>) (1999).
16. Melchiorri, A. *et al.* A measurement of Ω from the North American test flight of BOOMERanG. *Astrophys. J.* (submitted); also as preprint astro-ph/9911445 at (<http://xxx.lanl.gov>) (1999).
17. de Bernardis, P. *et al.* Mapping the CMB sky: the BOOMERanG experiment. *New Astron. Rev.* **43**, 289–296 (1999).
18. Mauskopf, P. *et al.* Composite infrared bolometers with Si_3N_4 micromesh absorbers. *Appl. Opt.* **36**, 765–771 (1997).
19. Bock, J. *et al.* Silicon nitride micromesh bolometer arrays for SPIRE. *Proc. SPIE* **3357**, 297–304 (1998).
20. Masi, S. *et al.* A self contained ^3He refrigerator suitable for long duration balloon experiments. *Cryogenics* **38**, 319–324 (1998).
21. Masi, S. *et al.* A long duration cryostat suitable for balloon borne photometry. *Cryogenics* **39**, 217–224 (1999).
22. Schlegel, D. J., Finkbeiner, D. P. & Davis, M. Maps of dust IR emission for use in estimation of reddening and CMBR foregrounds. *Astrophys. J.* **500**, 525–553 (1998).
23. Delabrouille, J., Gorski, K. M. & Hivon, E. Circular scans for CMB anisotropy observation and analysis. *Mon. Not. R. Astron. Soc.* **298**, 445–450 (1998).
24. Cheung, L. H. *et al.* 1.0 millimeter maps and radial density distributions of southern HII/molecular cloud complexes. *Astrophys. J.* **240**, 74–83 (1980).
25. Kogut, A. *et al.* Dipole anisotropy in the COBE DMR first-year sky maps. *Astrophys. J.* **419**, 1–6 (1993).
26. Tegmark, M. CMB mapping experiments: a designer’s guide. *Phys. Rev. D* **56**, 4514–4529 (1997).
27. Bond, J. R., Crittenden, R., Jaffe, A. H. & Knox, L. E. Computing challenges of the cosmic microwave background. *Comput. Sci. Eng.* **21**, 1–21 (1999).
28. Borrill, J. in *Proc. 3K Cosmology EC-TMR Conf.* (eds Langlois, D., Ansari, R. & Vittorio, N.) 277 (American Institute of Physics Conf. Proc. Vol. 476, Woodbury, New York, 1999).
29. Prunet, S. *et al.* in *Proc. Conf. Energy Density in the Universe* (eds Langlois, D., Ansari, R. & Bartlett, J.) (Editiones Frontieres, Paris, 2000).
30. Gorski, K. M., Hivon, E. & Wandelt, B. D. in *Proc. MPA/ESO Conf.* (eds Banday, A. J., Sheth, R. K. & Da Costa, L.) (European Southern Observatory, Garching); see also (<http://www.tac.dk/~healpix/>).
31. Kogut, A. in *Microwave Foregrounds* (eds de Oliveira Costa, A. & Tegmark, M.) 91–99 (Astron. Soc. Pacif. Conf. Series. Vol 181, San Francisco, 1999).
32. Toffolatti, L. *et al.* Extragalactic source counts and contributions to the anisotropies of the CMB. *Mon. Not. R. Astron. Soc.* **297**, 117–127 (1998).
33. Wright, A. E. *et al.* The Parkes-MIT-NRAO (PMN) surveys II. Source catalog for the southern survey. *Astrophys. J. Supp. Ser.* **91**, 111–308 (1994); see also (<http://astron.berkeley.edu/wombat/foregrounds/radio.html>).
34. Borrill, J. in *Proc. 5th European SGI/Cray MPP Workshop* (CINECA, Bologna, 1999); Preprint astro-ph/9911389 at (xxx.lanl.gov) (1999); see also (<http://cfpa.berkeley.edu/~borrill/cmb/madcap.html>).
35. Durrer, R., Kunz, M. & Melchiorri, A. *Phys. Rev. D* **59**, 1–26 (1999).
36. Seljak, U. & Zaldarriaga, M. A line of sight approach to cosmic microwave background anisotropies. *Astrophys. J.* **437**, 469–477 (1996).
37. Lewis, A., Challinor, A. & Lasenby, A. Efficient computation of CMB anisotropies in closed FRW models. Preprint astro-ph/9911177 at (<http://xxx.lanl.gov>) (1999).
38. Efstathiou, G. & Bond, R. Cosmic confusion: degeneracies among cosmological parameters derived from measurements of microwave background anisotropies. *Mon. Not. R. Astron. Soc.* **304**, 75–97 (1998).
39. Wright, E. *et al.* Comments on the statistical analysis of excess variance in the COBE-DMR maps. *Astrophys. J.* **420**, 1–8 (1994).
40. Perlmutter, S. *et al.* Measurements of Ω and Λ from 42 high-redshift supernovae. *Astrophys. J.* **517**, 565–586 (1999).
41. Schmidt, B. P. *et al.* The high-Z supernova search: measuring cosmic deceleration and global curvature of the Universe using type Ia supernovae. *Astrophys. J.* **507**, 46–63 (1998).

Acknowledgements

The Boomerang experiment was supported by Programma Nazionale di Ricerche in Antartide, Università di Roma “La Sapienza”, and Agenzia Spaziale Italiana in Italy, by the NSF and NASA in the USA, and by PPARC in the UK. We thank the staff of the National Scientific Ballooning Facility, and the United States Antarctic Program personnel in McMurdo for their preflight support and an effective LDB flight. DOE/NERSC provided the supercomputing facilities.

Correspondence and requests for materials should be addressed to P. d. B. (e-mail: debernardis@roma1.infn.it). Details of the experiment and numerical data sets are available at the web sites (<http://oberon.roma1.infn.it/boomerang>) and (<http://www.physics.ucsb.edu/~boomerang>).

Structure of the reovirus core at 3.6 Å resolution

Karin M. Reinisch*, Max L. Nibert† & Stephen C. Harrison*‡

* Harvard University and ‡ Howard Hughes Medical Institute, Fairchild Building, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA

† Department of Biochemistry and Institute for Molecular Virology, 1525 Linden Drive, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA

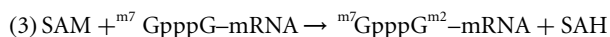
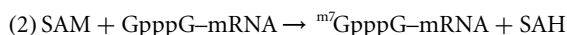
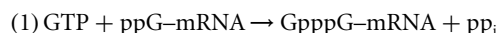
The reovirus core is an assembly with a relative molecular mass of 52 million that synthesizes, modifies and exports viral messenger RNA. Analysis of its structure by X-ray crystallography shows that there are alternative, specific and completely non-equivalent contacts made by several surfaces of two of its proteins; that the RNA capping and export apparatus is a hollow cylinder, which probably sequesters its substrate to ensure completion of the capping reactions; that the genomic double-stranded RNA is coiled into concentric layers within the particle; and that there is a protein shell that appears to be common to all groups of double-stranded RNA viruses.

Reoviruses, like other plant and animal double-stranded (ds) RNA viruses, contain a transcriptionally competent internal capsid particle (ICP), which remains intact after viral penetration¹. The reovirus ICP, known as the core, transcribes plus-strand copies from each of the ten genomic segments packaged within it, adds a methylated guanosine cap to the 5' end of each transcript and exports the mature mRNA into the cytoplasm of the infected cell. The core is thus an elaborately organized molecular machine.

About 700 Å in overall diameter, the reovirus core (Fig. 1) contains five of the eight proteins that make up a complete virion². Three of the five—λ1, λ2 and σ2—are symmetrically arranged in the icosahedral particle, and most of their amino-acid residues (4,500 in the icosahedral asymmetric unit) are clearly resolved in our refined structure. The reovirus core resembles in several respects the core of bluetongue virus (BTV)³, a member of a different genus of dsRNA viruses, but it has a number of distinctive features.

An unusual aspect of protein contacts in the reovirus core is the prevalence of non-equivalent interactions at many interfaces. The contacts within the λ1 shell, and even more strikingly those between this shell and the σ2 subunits that decorate it, show that when the pressure for genetic economy is strong enough, evolution can build at least two totally different specificities into a single protein surface. We draw similar conclusions from the published structure of the core of BTV³.

The sequence of modifications to the 5' end of the mRNA is incorporated into the spatial arrangement of enzymatic domains in the λ2 pentamer, a hollow cylinder that forms projecting turrets around each of the icosahedral fivefold axes. The reovirus mRNA cap consists of a 7-*N*-methyl guanosine linked by three phosphate groups to a 5'-terminal guanosine. The turrets catalyse the addition of this last guanosyl moiety to the mRNA as well as the transfer of a methyl group from *S*-adenosyl-*L*-methionine (SAM) to both the N7 of the added guanosine and to the 2' O of the first template-encoded nucleotide, which is also guanosine in all ten reovirus mRNAs:



where SAH is *S*-adenosyl-*L*-homocysteine⁴. Each of these reactions corresponds to a domain of λ2, which therefore resembles other complex enzymes that encode a biosynthetic pathway by concatenating domains to carry out successive reactions. The active

sites of the domains all face the interior of the pentameric λ2 turret, and the residence time of the mRNA 5' terminus within this hollow is probably sufficient to ensure completion of the cap.

Genomic dsRNA is coiled tightly within the reovirus core, in a manner reminiscent of the packing of dsDNA in phage heads⁵. The smooth inward-facing surface of the λ1 shell organizes the packed RNA into roughly concentric layers. Each segment of RNA is likely to be associated with a transcriptase complex tethered near a fivefold vertex^{6,7}.

The λ1 shell has structural relatives in other dsRNA viruses^{3,8} and it appears to be a common feature of nearly all viruses in this family. The other proteins of the ICP and especially those of the outer shell appear to vary among dsRNA viruses, corresponding to different strategies for infection and spread.

Structure determination

The very large unit cell (space group *F*432, *a* = 1,255 Å), the small crystal dimensions (about 150 μm) and the failure of cryopreservation

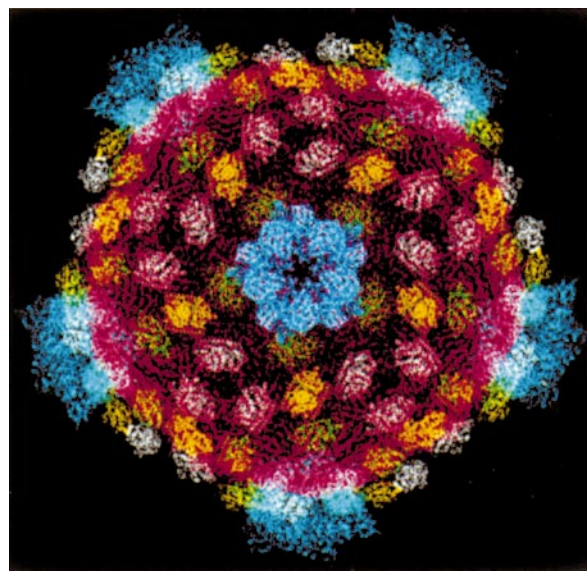


Figure 1 The reovirus core particle, represented by Cα traces of the constituent subunits. λ1 (relative molecular mass (*M*_r) 142K (ref. 46), 120 copies; shown in red) forms the shell that packages RNA and defines the symmetry and size of the particle. σ2 (*M*_r 47K, 150 copies; shown as yellow, green and white nodules) stabilizes the λ1 shell. λ2 (*M*_r 144K, 60 copies; shown in blue) forms turret-like structures around the fivefold axes that cap the nascent mRNA and organize its extrusion.

dictated our data-collection strategy. We determined phases at 27 Å resolution using a map reconstructed from electron cryomicroscopy (cryoEM)⁹ and extended using the fivefold non-crystallographic symmetry and solvent flattening.

The λ1 shell

The outer surface of the λ1 shell is relatively smooth, except for low ridges that border the binding sites of λ2 and σ2. Five monomers (set A) radiate from the fivefold axis like the petals of an inverted buttercup, and members of a further set of five (set B) interdigitate with the first. Twelve such decameric units together form a complete protein shell (Fig. 2).

Residues 240–1,275 of λ1 form a large, plate-like structure with no strongly defined domain boundaries (Fig. 3). Comparison of the A and B conformers reveals that two subdomains undergo a simple shift relative to one another: subdomain II (residues 482–922) is an insert into subdomain I (residues 253–470 and 923–1,260; Fig. 3). Each of the subdomains behaves as a rigid body when λ1A is transformed into λ1B, with minor local adjustments in loops at the perimeter of the subunit. Two of the peripheral loops in subdomain II—residues 560–568 and 774–794—block the opening through which plus-strand transcripts, destined to become mRNAs, are likely to pass from the central cavity into the λ2 turret. We believe that these loops will bend out of the way during active RNA synthesis. Direct evidence for their flexibility comes from inspection of λ1B, where residues 563–570 are disordered.

The two conformational states of λ1 allow it to fit into two distinct local environments. λ1A and B make similar contacts only along one side of subdomain I, where λ1A from one decamer faces λ1B from another decamer across a local dyad. This is the only instance of quasi-equivalence in the structure. The interactions of the two λ1 conformers, both with each other and with σ2, are otherwise very different (Fig. 2.).

The first 240 residues of λ1 are not part of the tightly folded, plate-like structure just described. Indeed, the amino-terminal residues are largely disordered, except for five copies per decamer (rather than ten) of each of three segments, containing residues 1–12, 40–167 and 181–240, respectively. The three ordered segments are part of a network of arms lying beneath the main part of the shell, and it is evident that the N-terminal residues of λ1 form variably disposed, extended structures that may help to tie the shell together. The way in which the ordered segments connect with each

other is not uniquely determined by well defined electron density, but the positions of the ends of these segments restrict possible connectivities, as summarized in Fig. 2c.

Residues 181–208, in the last of the three ordered arm segments just described, fold into a classical (Cys)₂(His)₂ zinc finger, as predicted from inspection of the sequence. Clear density for residues 181–208 and 215–240 and a lower-resolution bridge between them show that the ordered finger belongs to a λ1B subunit (Fig. 2c). The module is tucked into a pocket formed by three λ1 subunits. The N-terminal end of its α-helix, which contains the residues for specific DNA or RNA recognition in other zinc fingers, is accessible, but the surrounding parts of the λ1 shell would probably interfere with close approach of any genomic dsRNA segment. This finger would, however, be exposed at intermediate stages in assembly, and its variably positioned counterpart in the λ1A subunit is a candidate for RNA interaction, even in the mature core.

The σ2 clamp

The compact, globular monomer σ2 binds at three distinct locations within each icosahedral asymmetric unit (Figs 4 and 5). One of these sites (the position of σ2-i) lies over the middle part of λ1A. The second (σ2-ii) bridges from the middle of λ1B across to the carboxy-terminal portions of λ1B from another decamer. Although they receive the same face of σ2, these two sites are similar only in that each is a very shallow nest on the surface of the λ1 shell. Moreover, the interacting parts of σ2 undergo only modest conformational adjustments to fit tightly into these completely non-equivalent positions. The largest conformational difference is in residues 32–50, near the centre of the interface with the λ1 shell, where helix 39–46 unravels in σ2-ii. The third site for σ2 is directly across the icosahedral dyad over the C-terminal portions of λ1A. It contains a single σ2 (σ2-iii), in one of two equally likely, twofold-related orientations. The larger part of this interaction is identical to that of σ2-ii with λ1B.

The two unrelated σ2 binding sites are both of high affinity, as the σ2 subunit has only limited other contacts within the core but no tendency to dissociate. Its contacts with λ1, like those between neighbours in the λ1 shell, are marked examples of non-equivalent but strong and specific bonding, where a given surface of one subunit interacts with two completely different surfaces of another. When faced with the powerful pressures imposed by the compactness

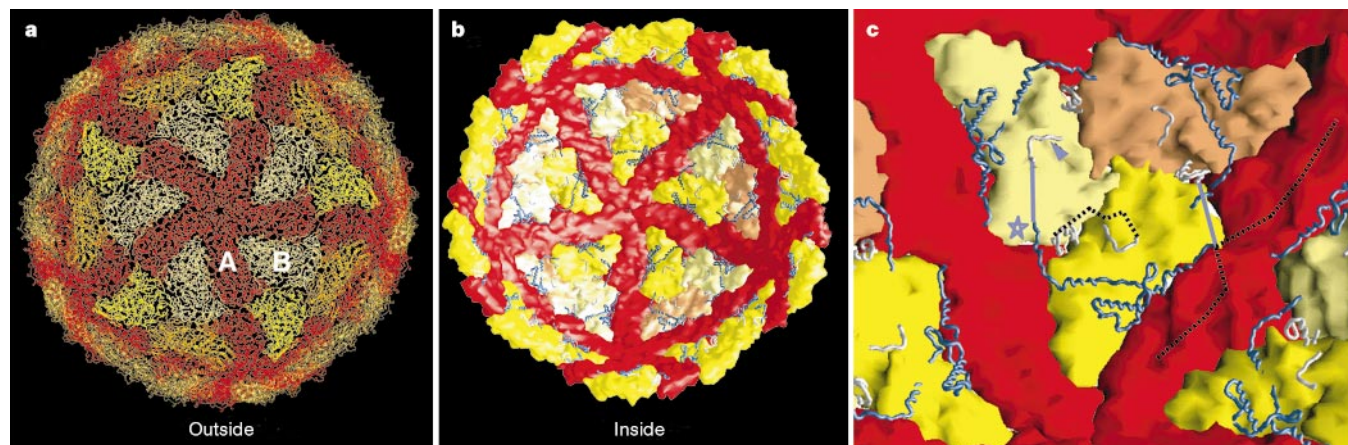


Figure 2 The λ1 shell. **a**, Cα trace, seen from the exterior. Five copies of λ1 (set A, in red) radiate from the icosahedral fivefold axis, and members of a second set (set B, in varying shades of yellow) interdigitate with the first to form a decamer. The interactions of the two λ1 conformers with each other are very different in all but one short interface across a local dyad. **b**, Schematic of the reovirus core from the interior. With the exception of three disconnected fragments, the first 240 residues of λ1 are disordered. These fragments include residues 1–12 (tentative assignment), 40–167 and 181–240. Residues 1–12,

40–167 and 181–208, a zinc finger, are shown. Residues 40–167 cannot belong to the λ1B subunit under which they lie, as the distance between 167 and 181 is too great. **c**, A small portion of **b**, magnified. A possible connectivity is shown in purple. In this scheme, residues 1–181 form an arm that ties together the λ1B subunits around a threefold axis. An arrow and a star mark the positions of residues 1 and 40, respectively. Dotted lines indicate other connectivities possible on the basis of distance criteria.

of viral genomes, proteins can evidently evolve to interact stably and specifically in two or more unrelated ways. The non-equivalent interdomain and intersubunit contacts in HIV reverse transcriptase¹⁰ appear to be the outcome of similar evolutionary ingenuity. In most virus structures analysed at high resolution, equivalent or quasi-equivalent interactions predominate. Even in polyoma¹¹ and SV40 (ref. 12), where 60 of the 72 VP1 pentamers are in 6- rather than 5-coordinated positions, use of arms to tie the structure together maximizes the local similarity of interpentamer contacts, despite the variable long-range geometry. As discussed below, interactions in the VP3 shell of BTV³ have the same non-equivalent characteristics as those in the λ 1 shell of reovirus, but there is no σ 2 equivalent. In the orbiviruses and other dsRNA viruses lacking a σ 2-like protein, the proteins that constitute the innermost shell (VP2 and VP3 in rota- and orbivirus, respectively) can self-assemble into icosahedral particles^{3,13,14}. When reovirus λ 1 is expressed in mouse L fibroblasts¹⁵ or in insect cells (M.L.N., unpublished data), however, no icosahedral particles form unless σ 2 is also expressed. The structure confirms the conclusion drawn from these results that σ 2 is a stabilizing clamp.

The λ 2 turrets are mRNA-capping complexes

The λ 2 pentamer is a hollow and slightly flared cylinder (Fig. 6). The cavity in its centre, through which mRNA must pass, is 70 Å in diameter at its widest, but it narrows to 15 Å at the top, where five flaps form a lid. The total volume of the channel is 2×10^5 Å³, which can accommodate up to about 300 nucleotides of nascent mRNA. The pentamers interact almost exclusively with λ 1A. Expressed alone λ 2 is monomeric^{16,17}, and it must therefore require a scaffold to assemble into turrets.

The λ 2 monomers have a series of seven domains, strung together with the most N-terminal domain at the base of the turret and the most C-terminal domain at the top (Fig. 6). The long axis of the subunit is at about 45° to the fivefold axis. The first 385 residues form a cup-like structure with its interior open into the hollow of the turret and its side resting on the λ 1 shell. Limited proteolysis of λ 2 has identified this domain as a guanylyltransferase (GTase). Analysis of a series of Lys-to-Ala mutations shows that Lys 190 forms the covalent nucleotidyl-enzyme intermediate and that Lys 171 also participates in GMP transfer¹⁷; both project into the active-site cavity (Fig. 6d). A narrow channel gives access to the active site for nucleotides from outside the turret. The structure of

the reovirus GTase appears to be a new fold; it differs from the known structure of PBCV-1 GTase¹⁸.

Residues 386–433 and 690–802 contribute to a small domain that bridges the GTase and the methylases. It is a four-strand sheet, protected by α -helices facing toward the outside of the turret, and it probably functions as a structural support for the three enzyme domains.

Residues 434–691 form a methyltransferase domain (methylase-1). Helices sandwich a seven-strand β -sheet of mixed polarity in a variant of the SAM-binding domain seen in numerous other methyltransferases (Fig. 6)^{19,20}. The SAM site is conserved among

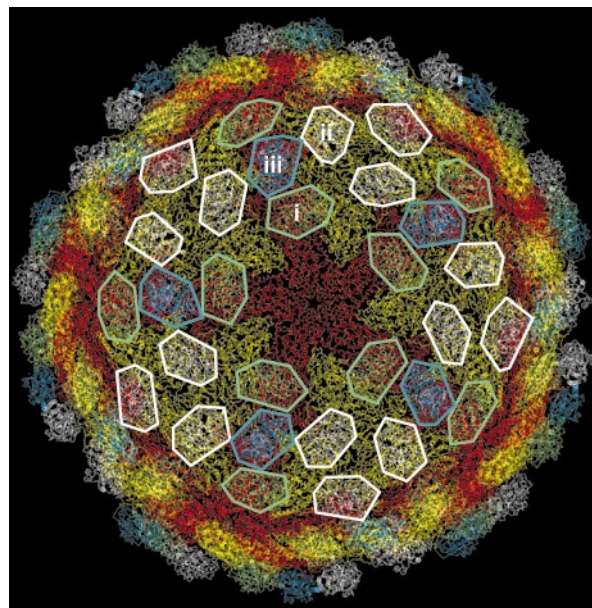


Figure 4 Subunit σ 2 has 150 binding positions on the λ 1 surface. There are 60 copies each of sites i (green outlines) and ii (white outlines) and 30 of site iii (blue outlines). Site iii lies on an icosahedral dyad, and σ 2 in that position adopts two opposite orientations in 50:50 ratio. Sites i and ii involve very different residues of λ 1; each identical half of site iii (on λ 1A) is the same as part of site ii (on λ 1B). Most of the contacts of σ 2 are with the λ 1 shell, but σ 2-iii has sparse contacts with both σ 2-i and ii, and σ 2-i has a modest contact with λ 2.

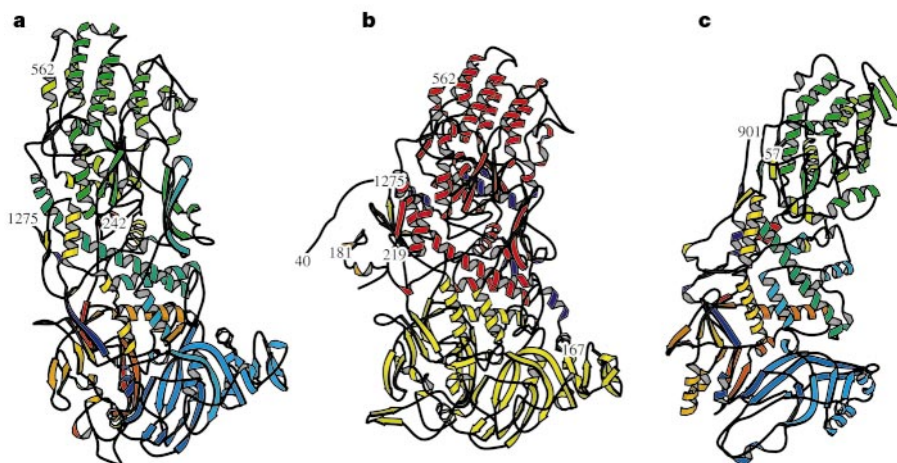


Figure 3 Two conformations of λ 1 and comparison with BTV VP3. **a**, Course of the polypeptide chain of λ 1A, in colours graded from red at the N terminus to blue at the C terminus. λ 1 is mostly α -helical, with significant regions of β -sheet present only in the part furthest from the fivefold axis. **b**, λ 1B; subdomains I and II are yellow and red, respectively. λ 1A and B differ with respect to a conformational shift about a pivot between

the two subdomains. The zinc finger (residues 181–208) is gold, and residues 40–167 from a neighbouring λ 1 subunit are blue. **c**, VP3-A, the λ 1A equivalent from BTV, in colours graded from red at the N terminus to blue at the C terminus³. Although λ 1 and VP3 have no sequence similarity and details of their folds differ, their overall conformations are clearly closely related.

these enzymes, and we can thus predict where SAM will bind to methylase-1 (Fig. 6). Difference maps computed from diffraction data taken after diffusion of SAH into the core crystals confirm that SAH binds in this location, in a channel that connects the interior and the exterior of the turret. A conformational change in residues 519–524 and 579–587, such that the 580s loop blocks off the exterior of the channel, accompanies SAH binding. The mRNA cap must bind close by, in the cavity between methylase-1 and the GTase domain of an adjacent $\lambda 2$ monomer. Sharing of the active site between two monomers explains the absence of methylase activity in monomeric $\lambda 2$ (ref. 16).

Residues 804–1,022 form the second methylase domain (methylase-2, Fig. 6). The order and directionality of the strands and the positions of some α -helices are as in the consensus fold; residues 926–962 and 983–1,017 represent an insertion, forming a three-strand parallel sheet and two helices. The SAM-binding site, identified from the diffusion experiment described above, lies as anticipated from homologous enzymes in another channel to the exterior of the turret. SAH binding does not induce a conformational change at this site.

tional change at this site.

The last 250 residues of the $\lambda 2$ polypeptide chain form a three-domain flap at the top of the turret. In virions and intermediate subviral particles (ISVPs), this flap is a cantilever to anchor the cell-attachment protein, $\sigma 1$ (ref. 9). In cores, it may serve as a gate to retard exit of the 5' terminus of the mRNA. The three domains all have structures related to the immunoglobulin fold. The first two resemble the V and C regions of an antibody light chain, although they lack disulphide bonds; the third is a sandwich of two three-strand sheets, folded like a truncated V domain.

Which domain is the 7'-N- and which the 2'-O-methylase? And what determines the order of the modification reactions? The view of two adjacent monomers in Fig. 6f shows that although there is no groove to guide the mRNA terminus from one active site to another, the spatial arrangement of sites in the turret may reflect the temporal sequence of the capping reactions. Guanylyl transfer occurs first, near the base of the turret, where newly synthesized transcript enters the multi-enzyme cavity. The GTase site from one monomer (for example, monomer A in Fig. 6f) is closer to the

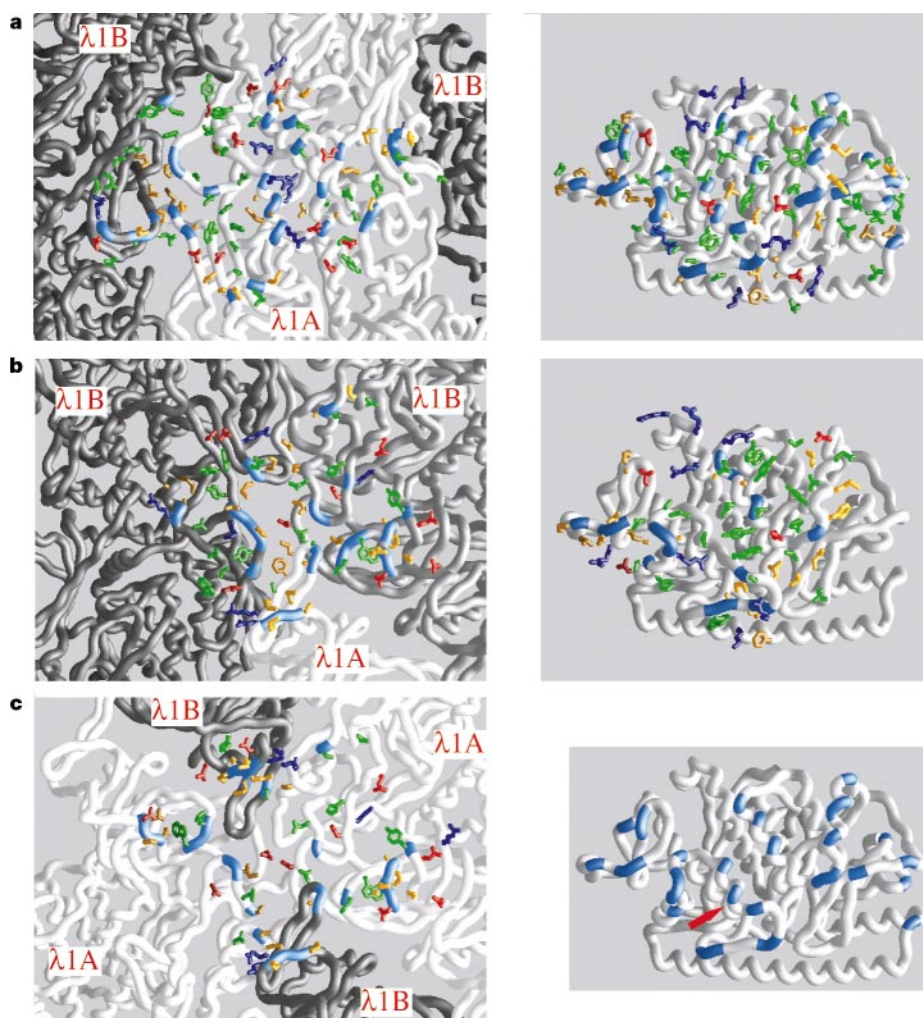


Figure 5 The $\lambda 1$ – $\sigma 2$ interface. **a, b**, Left, the $\lambda 1$ surfaces onto which bind $\sigma 2$ -i and $\sigma 2$ -ii, respectively. Right, the corresponding contacting face of $\sigma 2$ -i and $\sigma 2$ -ii. The $\lambda 1$ surfaces are drawn to receive $\sigma 2$ -i and $\sigma 2$ -ii in the same orientation; the images of $\sigma 2$ on the right may be flipped around the axis parallel to the long side of the page to orientate them correctly over the corresponding images of $\lambda 1$ on the left. **c**, Left, site iii on the $\lambda 1$ surface. Right, $\sigma 2$ -i, with a red arrow indicating an α -helix that unravels in $\sigma 2$ -ii. Different $\lambda 1$ subunits are painted in different shades of grey except in **c**, where both $\lambda 1$ As are white and both $\lambda 1$ Bs are dark grey. Main-chain segments within 4 Å of the interacting surface

are coloured light blue. Side chains that fall within 4 Å of the interacting surface are blue (positively charged), red (negatively charged), green (neutral polar) or orange (hydrophobic). $\sigma 2$ -i binds mostly to $\lambda 1$ A, whereas $\sigma 2$ -ii binds mostly to two copies of $\lambda 1$ B and acts as a clamp. The zinc finger (not shown) lies inside the core under $\sigma 2$ -ii but makes no contacts with it. $\sigma 2$ -iii binds at the icosahedral twofold axis; note the similarity of the $\lambda 1$ surface with which it interacts (left) to the white and light grey portions in **b**. $\lambda 1$ residues marked in the last panel on the left are inferred, as $\sigma 2$ -iii has not been built explicitly into the density.

methylase-1 active site in the clockwise neighbour (monomer B in Fig. 6f) than to any other methylase-1 site, so the newly guanylated 5' end is more likely to pass from the GTase on one subunit to methylase-1 on the neighbour than to any other methylase site. As N7 methylation precedes 2'O methylation⁴, we suggest that methylase-1 is the N7-modifying enzyme and that methylase-2 is the O-methyltransferase. In this tentative assignment, the sequence of domains in the subunit corresponds to the order of reactions in the pathway. N7 methylation would take place in the middle of the turret, and the last reaction, 2'O methylation, would take place at the top, near the exit pore. It is reasonable to suppose that the 2'O methylase, like the related enzyme from vaccinia virus²¹, accepts

termini only when the base of the 5'-guanylyl group has been methylated at N7.

Containment, to allow the 5' end of the transcript to visit various sites, perhaps nearly randomly, but also to ensure that modification will be complete before the cap emerges into the cytoplasm, appears to be a principle embodied in the design of the λ 2 turret. The turret is a container with several, densely arrayed, inward-facing active sites. Similar arrangements have been seen in chaperonins such as GroEL and in intracellular proteases such as ClpAP and the 26S proteasome (see ref. 22 and references therein). In GroEL, release of substrate is timed by ATP hydrolysis; in λ 2, release of the transcript 5' terminus is more likely to be triggered simply by filling of the

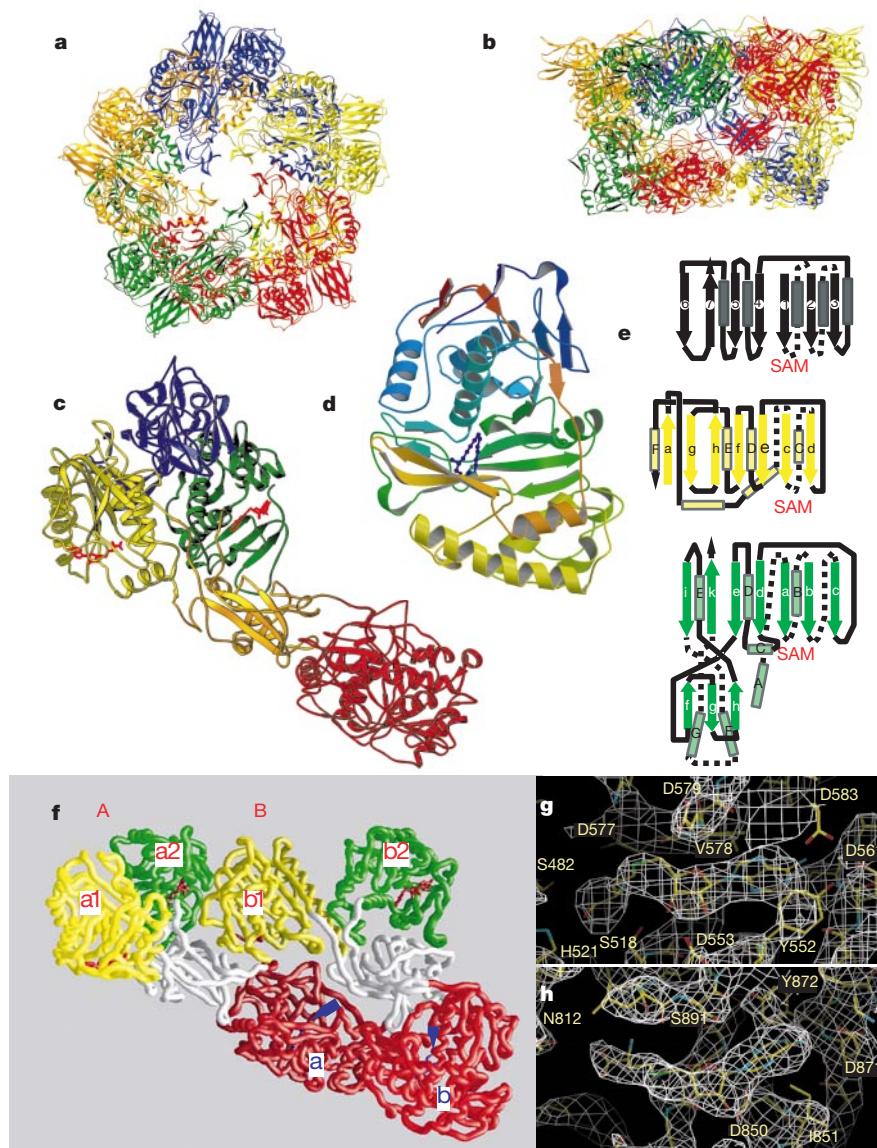


Figure 6 The capping complex. **a**, **b**, λ 2 turret (about 120 Å diameter and 80 Å tall) viewed from the top and side, respectively. The five elongated λ 2 monomers, each shown in a different colour, wrap around the outer surface, with their long axes at about 45° to the radial direction. **c**, The λ 2 monomer, viewed from the inside of the pentamer (the blue monomer in **b**). The GTase domain is red, methylase-1 is yellow, methylase-2 is green and the immunoglobulin-like domains are blue. Red SAH molecules mark the SAM-binding sites. **d**, GTase domain at 90° to **c**, in graded colours with the N terminus red and the C terminus blue. Side chains of K190 and K171 are shown. **e**, Diagrams of SAM-binding domains. The 'universal' SAM-binding domain is shown in black^{19,20}; methylase-1 is yellow and methylase-2 is green. Secondary structural elements are aligned vertically

with their equivalents in the universal fold. The SAM-binding position with respect to the β -sheet is labelled. **f**, Two monomers of λ 2 (labelled A and B) with the immunoglobulin-like domains detached and viewed from the interior of the turret. Monomer A includes GTase 'a' and methylases 'a1' and 'a2'; B includes 'b', 'b1' and 'b2'. Coloured as in **c**. Red SAH molecules indicate SAM-binding locations. Blue arrows indicate the GTase active site. **g**, SAH density for methylase-1 in a 4 Å, $2F_o - F_c$ map made with data from crystals soaked in 2 mM SAH. SAH binding is accompanied by conformational changes in residues 519–524 and 579–587. Some residues that may interact with SAH are labelled. **h**, SAH density for methylase-2.

cavity. Capping enzyme complexes that have only a single copy of each capping enzyme and that do not sequester the nascent RNA within a λ 2-like container may have some other guiding mechanism to order the progress of the 5' terminus from one active site to the next.

Interior of the core

The inward-facing surface of the λ 1 shell defines a cavity with a mean outer radius of 245 Å. The N-terminal extensions of λ 1 are the most significant irregularities in the otherwise smooth cavity boundary. Reconstructed cryoEM images of empty virions and cores show, however, that there are structures projecting inwards near the fivefold axes—presumably the transcriptase complexes⁶. The presence of a single such complex near each fivefold axis will leave little trace in a high-resolution map.

When packaged within a hollow shell, double-helical nucleic acids form uniform coils that minimize kinks and sharp curves while maximizing the interhelix distance (and hence minimizing electrostatic repulsion)^{5,23–27}. Assuming that the entire interior of the core (after allowing for the volume occupied by non-icosahedrally ordered protein) is uniformly occupied by RNA with local hexagonal packing, the mean spacing between adjacent segments of A-form dsRNA will be about 30 Å. This figure agrees with the spacing predicted from the 26 Å diffraction maximum seen in solution X-ray scattering from intact type 3 reovirus particles²⁴. A similar observation has been reported for BTV cores⁷.

Although we do not see a strong 26 Å diffraction ring, in low-resolution maps we observe three or four shells of density, spaced at 26 Å and layered inwards from the interior surface of λ 1. The shells are weak in the region beneath each fivefold vertex and stronger elsewhere. We interpret these observations, together with those from earlier experiments²⁴, as indicating that dsRNA is coiled within the shell. We expect that an outer layer of RNA will press against λ 1 and that subsequent layers will derive their approximate coherence from the balance of forces described above. Thus, the RNA exhibits significant statistical order, but there is no evidence for unique three-dimensional packing. The presence of a transcriptase complex tethered near each fivefold vertex will tend to exclude RNA from the vicinity of the fivefold axes, and these complexes may even act as spools for coiling individual RNA segments²⁸. Thus, we expect the shells of density to be weak beneath each turret, as observed.

Comparisons with other dsRNA viruses

The transcriptionally competent ICPs from members of the nine known genera of plant and animal dsRNA viruses in the family *Reoviridae* all appear to share a common, λ 1-like framework, which packages the genome and organizes the polymerization and capping enzymes. The organization of this shell may extend to fungal and bacterial viruses with dsRNA genomes^{3,28,29}. The equivalent framework in the orbivirus core, which has been seen at high resolution³, consists of 120 copies of VP3, which is arranged identically to λ 1. Moreover, λ 1 (residues 240–1,275) and VP3 have the same plate-like shape, and although they have no sequence similarity, they have similar overall folds with corresponding parts of their polypeptide chains forming corresponding portions of each molecule. In each protein, an α -helical bundle lies closest to the icosahedral fivefold axis, a β -sheet 'domain' lies most distally from that axis, and a lattice of criss-crossing α -helices lies in the middle. The details of the fold differ, however, and it is not possible meaningfully to superimpose any significant substructure of λ 1 onto VP3. Furthermore, whereas the two conformers of λ 1 differ in the relative orientations of subdomains I and II about one pivot, the two conformers of VP3 differ in the relative orientations of three subdomains about two such pivots, neither of which corresponds to the λ 1 pivot.

Beyond the interior framework, the structures of the dsRNA viruses diverge, reflecting different mechanisms for viral maturation and entry. The mammalian orthoreoviruses have turreted ICPs: the capping complexes (λ 2) are ordered structures displayed on the outside of the λ 1 shell. There are also clamp subunits (σ 2) to reinforce λ 1 interactions. Members of four other dsRNA genera also have turret-like structures that are likely to be homologues of λ 2. Two of these genera, the aquareoviruses³⁰ and the cypoviruses^{8,31}, have been studied by cryoEM. They have nodules at positions corresponding to σ 2-i and σ 2-ii, where the protein at position ii is probably a clamp required for assembly⁸. In contrast, the orbiviruses (and two other genera) have non-turreted ICPs coated with a complete $T = 13$ lattice of a non-enzymatic protein (VP7 and VP6, respectively)^{3,32}, and their capping enzymes lie inside the common shell, perhaps in close association with the polymerases.

Discussion

We draw four conclusions from the above data. First, there is a

Table 1 Crystallographic statistics

Native data					
Resolution range (Å)	R_{sym} (%)	R_{cum} (%)	Completion (%)	Redundancy	I/σ
70–9.82	8.9	8.9	99.4	6.8	6.5
9.82–6.95	13.1	11.1	99.7	6.6	5.3
6.95–5.67	25.0	14.9	99.7	6.3	2.9
5.67–4.91	25.9	17.9	99.7	6.2	2.9
4.91–4.39	25.7	20.0	99.6	6.0	2.9
4.39–4.01	35.7	22.9	99.4	5.6	2.1
4.01–3.71	47.9	25.3	97.4	4.1	1.6
3.71–3.47	53.0	26.0	70.5	2.1	1.4
SAH soaking data					
Resolution range (Å)	R_{sym} (%)	R_{cum} (%)	Completion (%)	Redundancy	I/σ
70–10.44	7.4	7.4	57.7	1.5	8.3
10.44–7.38	11.5	9.7	57.7	1.5	6.4
7.38–6.03	28.0	14.4	56.9	1.5	2.7
6.03–5.22	33.8	18.8	55.3	1.5	2.2
5.22–4.67	31.7	21.6	51.2	1.4	2.3
4.67–4.26	35.7	23.9	45.0	1.4	2.1
4.26–3.95	49.7	26.1	37.7	1.3	1.5
Refinement: R -value for working set					
Resolution range (Å)	20–3.6	(3.64–3.6)			
R (%)	20.81	(30.06)			
No. of reflections	831,984	(23,226)			

$R_{\text{sym}} = \sum_i \sum_j |I_i(h) - \langle I(h) \rangle| / \sum_i \sum_j I_i(h)$, where $I_i(h)$ is the i th measurement of reflection h and $\langle I(h) \rangle$ is the weighted mean of all measurements of h . $R = \sum_n |F_{\text{obs}}(h) - k|F_{\text{calc}}(h)| / \sum_n |F_{\text{obs}}(h)|$, where k is a scale factor.

striking and unexpected degree of non-equivalence in the inter-subunit contacts that stabilize the shell. Second, the RNA-capping complex contains redundant copies of each active site, facing into the cavity of a hollow cylinder. Temporary retention of the 5' terminus within the cavity probably ensures efficient capping. Third, genomic RNA is tightly packed within the interior of the core, leading to a statistical ordering. Repeated passage of each genome segment past a polymerase tethered near the exit point for the transcript will require a concerted set of activities and motions. Finally, known dsRNA viruses are constructed around a conserved framework, the homologue of $\lambda 1$, with considerable variability in the way other viral proteins, which function in entry and exit, associate with the common shell. □

Methods

We prepared cores of reovirus strain F18 as described^{2,33}. For crystallization, cores (6 mg ml⁻¹ in 0.5 M NaCl, 0–0.3 M NaOAc, 50 mM MgCl₂, 50 mM Hepes, pH 7) were equilibrated against 1 M NaCl, 0–0.6 M NaOAc, 100 mM HEPES, pH 7.0. Crystals, 150–200 μ m at their largest dimensions, grew in 9–12 months. They were transferred to a solution containing 0.8 M NaOAc (but otherwise identical to the well solution) and mounted in capillaries. The space group is *F*432, *a* = 1,255 Å. Data were collected at 0 °C at CHESS F1. One 0.12° wedge of data was collected from each crystal volume, defined by a 0.1 mm collimator, during an exposure of 40 s. We recorded data between 65 and <3.6 Å on two Fuji image plates mounted in a specially constructed double-cassette holder placed 700 mm from the crystal. More than 500 such exposures contributed to the final data set; many others were discarded. Each double image was indexed and integrated with Denzo³⁴; intensities were scaled using Scala and Postref in the CCP4 program suite³⁵. Reflections more than 75% complete were scaled up. Statistics are shown in Table 1. The merged data were sharpened with a *B* factor of –55.

We used a 27 Å cryoEM reconstruction of the core⁹ as a phasing start. To minimize the effect of contrast-transfer-function errors, we converted the cryoEM map into a binary map to represent protein/non-protein regions. We used MAVER (RAVE software package³⁶) to position the model in the *F*432 asymmetric unit. The translation and orientation of each particle was fixed by packing and space group. Each particle was centred at (1/4, 1/4, 1/4) and related positions, with icosahedral two- and threefold axes aligned with crystallographic ones. There are two such orientations, and they correspond to a choice of origin. One pentameric moiety of the icosahedral core lies in each *F*432 asymmetric unit. We adjusted the magnification of the model to maximize correlation and to minimize *R* factor between *F*_{calc} and *F*_{obs} in a resolution range 65–27 Å (*R* = 46.5%, Corr = 59.8%). RAVE was used both to construct an envelope³⁷ and, with CCP4, for all averaging and phase extension. Minor modifications of RAVE and CCP4 were necessary. We carried out phase extension in steps of one reciprocal lattice point. Unmeasured reflections, including all at resolutions lower than 65 Å, were substituted with calculated values, and weights for the terms in *2F*_o–*F*_c maps were assigned in SIGMAA. To extend from 4 to 3.6 Å, some portions of the asymmetric unit ($\sigma 2$ -i and ii, and subdomains I and II of $\lambda 1$) were averaged tenfold. The resulting maps were unambiguous, and models for $\lambda 2$ and two versions of $\lambda 1$ and $\sigma 2$ were built using program O³⁸; a partial model of $\sigma 2$ -ii was docked into $\sigma 2$ -iii density during refinement. We determined Pt sites from 1.5° of an 8 Å data set obtained from crystals soaked in 1 mM of K₂PtCl₄; they correspond to methionine and histidine positions in the structure. Positional refinement, torsion angle dynamics and restrained individual *B*-factor refinement were carried out in CNS³⁹, maintaining strict fivefold restraints. To speed calculations, we carried out refinement in a primitive cell with smaller unit-cell dimensions (*a* = *b* = *c* = 887.42 Å, α = β = γ = 60°) rather than the face-centred cell³⁹ (R. Grosse-Kunstleve, personal communication). Portions of the structure were also refined in small P1 cells using artificial '*F*_{obs}' calculated from portions of an averaged *2F*_o–*F*_c map in *F*432 (ref. 40). The cross-validation set was not used in calculating these averaged maps. The final *R* factor is 20.8% (*R*_{free} = 20.6%; Table 1). The r.m.s. deviation from ideal bond lengths is 0.009 Å; that from ideal angles is 1.3°. Of all residues, 99.4% fall in either the most favoured or allowed region of the Ramachandran plot, and only 3 out of 4,733 fall in the disallowed region⁴¹.

To determine SAH-binding sites, we soaked crystals in mounting buffer supplemented with 2 mM SAH for several days. A partial data set (60–4 Å) was collected; *2F*_o–*F*_c and *F*_o–*F*_c difference maps were calculated and averaged fivefold using a liberal envelope. Strong density appeared in the SAM-binding locations. Difference density also clearly indicated a backbone movement for residues 579–587 and 519–524 in the methylase-1 domain.

We computed low-resolution maps for visualizing RNA density using additional data between 200 and 50 Å collected on BioCARS beamline 14-ID-B at the Advanced Photon Source. Phase extensions starting at 33 Å and at 27 Å and using a variety of models and final resolution limits yielded similar results: density ripples spaced at 25–26 Å intervals from the inner surface of the core and interrupted under the fivefold axes.

Figures were prepared in Ribbons⁴², Molscript⁴³, Raster3D⁴⁴, GRASP⁴⁵ and O³⁸.

Received 6 January; accepted 15 March 2000.

- Fields, B. N. in *Fields Virology* (eds Fields, B. N., Knipe, D. M. & Howley, P. M.) 1553–1555 (Lippincott-Raven, Philadelphia, 1996).
- Nibert, M. L., Schiff, L. A. & Fields, B. N. in *Fields Virology* (eds Fields, B. N., Knipe, D. M. & Howley, P. M.) 1557–1596 (Lippincott-Raven, Philadelphia, 1996).

- Grimes, J. M. *et al.* The atomic structure of the bluetongue virus core. *Nature* **395**, 470–478 (1998).
- Furuichi, Y., Muthukrishnan, S., Tomasz, J. & Shatkin, A. Mechanism and formation of reovirus mRNA 5'-terminal blocked and methylated sequence ^{m7}GpppG^{m7}pC. *J. Biol. Chem.* **251**, 5043–5053 (1976).
- Earnshaw, W. C. & Harrison, S. C. DNA arrangement in isometric phage heads. *Nature* **268**, 598–602 (1977).
- Dryden, K. A. *et al.* Internal structures containing transcriptase-related proteins in top component particles of mammalian orthoreovirus. *Virology* **245**, 33–46 (1998).
- Gouet, P. *et al.* The highly ordered double-stranded RNA genome of bluetongue virus revealed by crystallography. *Cell* **97**, 481–490 (1999).
- Hill, C. L. *et al.* The structure of cypovirus and the functional organization of dsRNA viruses. *Nature Struct. Biol.* **6**, 565–568 (1999).
- Dryden, K. A. *et al.* Early steps in reovirus infection are associated with dramatic changes in supramolecular structure and protein conformation: analysis of virions and subviral particles by cryoelectron microscopy and image reconstruction. *J. Cell Biol.* **122**, 1023–1041 (1993).
- Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. Crystal structure of 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256**, 1783–1790 (1992).
- Stehle, T., Yan, Y., Benjamin, T. L. & Harrison, S. C. Structure of murine polyomavirus complexed with an oligosaccharide receptor fragment. *Nature* **369**, 160–163 (1994).
- Liddington, R. C. *et al.* Structure of simian virus 40 at 3.8 Å resolution. *Nature* **354**, 278–284 (1991).
- Labbe, M., Charpilienne, A., Crawford, S. E., Estes, M. K. & Cohen, J. Expression of rotavirus VP2 produces empty corelike particles. *J. Virol.* **65**, 2946–2952 (1991).
- Moss, S. R. & Nuttall, P. A. Subcore- and core-like particles of Broadhaven virus (BRDV), a tick-borne orbivirus, synthesized from baculovirus expressed VP2 and VP7, the major core proteins of BRDV. *Virus Res.* **32**, 401–407 (1994).
- Xu, P., Miller, S. & Joklik, W. K. Generation of reovirus core-like particles in cells infected with hybrid vaccinia viruses that express genome segments L1, L2, L3, and S2. *Virology* **197**, 726–731 (1993).
- Mao, Z. & Joklik, W. K. Isolation and enzymatic characterization of protein $\lambda 2$, the reovirus guanylyltransferase. *Virology* **185**, 377–386 (1991).
- Luongo, C. L., Reinisch, K. M., Harrison, S. C. & Nibert, M. L. Identification of the guanylyltransferase region and active site in reovirus mRNA capping protein 12. *J. Biol. Chem.* **275**, 2804–2810 (2000).
- Håkansson, K., Doherty, A. J., Shuman, S. & Wigley, D. B. X-ray crystallography reveals a large conformational change during guanylyl transfer by mRNA capping enzymes. *Cell* **89**, 545–553 (1997).
- Schluckebier, G., O'Gara, M., Saenger, W. & Cheng, X. Universal catalytic domain structure of AdoMet-dependent methyltransferases. *J. Mol. Biol.* **247**, 16–20 (1995).
- Hodel, A. E., Gershon, P. D., Shi, X. & Quijcho, F. A. The 1.85 Å structure of the vaccinia protein VP39: a bifunctional enzyme that participates in the modification of both mRNA ends. *Cell* **85**, 247–256 (1996).
- Hu, G., Gershon, P. D., Hodel, A. E. & Quijcho, F. A. mRNA cap recognition: dominant role of enhanced stacking interactions between methylated bases and protein aromatic side chains. *Proc. Natl Acad. Sci. USA* **96**, 7149–7154 (1999).
- Wickner, S., Maurizi, M. R. & Gottesman, S. Posttranslational quality control: Folding, refolding, and degrading proteins. *Science* **286**, 1888–1893 (1999).
- Earnshaw, W. C., King, J., Harrison, S. C. & Eiserling, F. A. The structural organization of DNA packaged within the heads of T4 wild-type, isometric and giant bacteriophages. *Cell* **14**, 559–568 (1978).
- Harvey, J. D., Bellamy, A. R., Earnshaw, W. C. & Schutt, C. Biophysical studies of reovirus type 3: iv low-angle x-ray diffraction studies. *Virology* **112**, 240–249 (1981).
- Harrison, S. C. Packaging of DNA into bacteriophage heads: a model. *J. Mol. Biol.* **171**, 577–580 (1983).
- Cerritelli, M. E. *et al.* Encapsidated conformations of bacteriophage T7 DNA. *Cell* **91**, 271–280 (1997).
- Booy, F. P. *et al.* Liquid crystalline, phage-like packing of encapsidated DNA in herpes simplex virus. *Cell* **64**, 1007–1015 (1991).
- Butcher, S. J., Dokland, T., Ojala, P. M., Bamford, D. H. & Fuller, S. D. Intermediates in the assembly pathway of the double-stranded RNA virus $\phi 6$. *EMBO J.* **16**, 4477–4487 (1997).
- Cheng, R. H. *et al.* Fungal virus capsids, cytoplasmic compartments for the replication of double-stranded RNA, formed as icosahedral shells of asymmetric Gag dimers. *J. Mol. Biol.* **244**, 255–258 (1994).
- Shaw, A. L., Samal, S. K., Subramanian, K. & Prasad, B. V. V. The structure of aquareovirus shows how the different geometries of the two layers of capsid are reconciled to provide symmetrical interactions and stabilization. *Structure* **4**, 957–967 (1996).
- Zhang, H. *et al.* Visualization of protein–RNA interactions in cytoplasmic polyhedrosis virus. *J. Virol.* **73**, 1624–1629 (1999).
- Prasad, B. V., Wang, G. J., Clerx, J. P. & Chiu, W. Three-dimensional structure of rotavirus. *J. Mol. Biol.* **199**, 269–275 (1988).
- Coombs, K. M., Fields, B. N. & Harrison, S. C. Crystallization of the reovirus type 3 Dearing core. *J. Mol. Biol.* **215**, 1–5 (1990).
- Otwiński, Z. & Minor, W. in *Macromolecular Crystallography A* (eds Carter, C. W. & Sweet, R. M.) 307–326 (Academic, New York, 1997).
- CCP4. The CCP4 suite: programs for X-ray crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Jones, T. A. in *Molecular Replacement* (eds Dodson, E. J., Gover, S. & Wolf, W.) 91–105 (SERC Daresbury Laboratory, Warrington, 1992).
- Kleywegt, G. J. & Jones, T. A. in *From First Map to Final Model* (eds Bailey, S., Hubbard, R. & Waller, D.) 59–66 (SERC Daresbury Laboratory, Warrington, 1994).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brunker, A. T. *et al.* Crystallography & NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Jacobsen, D. H., Hogle, J. M. & Filman, D. J. A pseudo-cell based approach to efficient crystallographic refinement of viruses. *Acta Crystallogr. D* **52**, 693–711 (1996).
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).

42. Carson, M. in *Macromolecular Enzymology* (eds Carter, C. W. Jr & Sweet, R. M.) 493–505 (Academic, New York, 1996).
43. Esnouf, R. M. An extensively modified version of Molscript that includes greatly enhanced coloring capabilities. *J. Mol. Graphics* **15**, 133–138 (1997).
44. Merrit, E. A. & Murphy, M. E. P. Raster3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).
45. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).
46. Harrison, S. J. *et al.* Mammalian reovirus L3 gene sequences and evidence for a distinct amino-terminal region of the lambda1 protein. *Virology* **258**, 54–64 (1999).

Acknowledgements

We thank T. F. Severson, R. L. Margraf and S. J. Harrison for growing reovirus-infected cells; T. S. Baker and S. B. Walker for providing us with an unpublished cryoEM reconstruction of cores from reovirus reassortant F18; D. Thiel and the staff of CHESS and

McCHESS and R. Pahl and the staff of the BioCars beamline 14-ID-B for assistance; members of the Harrison laboratory, especially J. Cruzan, for scanning image plates; W. Minor for help with data collected with an early version of the double image plate cassette holder; B. Miller for constructing an improved vacuum-based version; S. Ray for advice on computing strategies; P. Adams and R. Grosse-Kunstleve for advice on refinement; and R. Grosse-Kunstleve for implementing refinement in the smaller rhombohedral cell. K.M.R. thanks D. W. Rodgers for help in the initial cryopreservation attempts, for suggesting a vacuum-based design for the double image plate cassette holder and for advice. We acknowledge support from the NIH (to S.C.H. and M.L.N.). S.C.H. is an investigator in the Howard Hughes Medical Institute. M.L.N. receives support from a grant of the Lucille P. Markey Charitable Trust to the Institute for Molecular Virology and, as a Shaw Scientist, from the Milwaukee Foundation. This work was started with the stimulus and collaboration of the late B. N. Fields.

Correspondence and requests for material should be addressed to S.C.H. (e-mail: harrison@crystal.harvard.edu). The atomic coordinates have been deposited in the RCSB Protein Data Bank under accession number 1EJ6.

Identification of molecular-cloud material in interplanetary dust particles

Scott Messenger

McDonnell Center for the Space Sciences, Physics Department,
Washington University, St. Louis, Missouri 63130-4899, USA

Interplanetary dust particles (IDPs) collected in the Earth's stratosphere and meteorites are fragments of comets and asteroids. These are 'primitive' meteorites in part because they have preserved materials which predate the formation of the Solar System. The most primitive (least altered) meteorites contain a few parts per million of micrometre-sized dust which formed in the atmospheres of giant stars¹. Some meteorites² have elevated D/H and ¹⁵N/¹⁴N ratios that are attributed to surviving interstellar organic molecules which have probably been strongly diluted and altered by parent-body processes². Most IDPs are chemically, mineralogically, and texturally primitive in comparison to meteorites^{3,4}. Here I show that H and N isotopic anomalies among fragile 'cluster' IDPs are far larger, more common, and less equilibrated than those previously observed in other IDPs or meteorites. In some cases, the D/H ratios that we measure reach the values of interstellar molecules, suggesting that molecular-cloud material has survived intact. These observations indicate that cluster IDPs are the most primitive class of Solar System materials currently available for laboratory analysis.

Interplanetary dust particles are a diverse assemblage of materials

which are commonly very fine grained ($\sim 0.1 \mu\text{m}$), heterogeneous and rich in volatile elements and carbon relative to all meteorite classes^{5,6}. Hydrogen and nitrogen isotopic anomalies have been previously observed among IDPs, although with only one exception these isotopic effects have not exceeded the values observed among the most primitive meteorites⁷⁻¹⁰. Here I extend these previous studies to cluster IDPs—particles which are so fragile that they disintegrate into dozens or hundreds of fragments upon impact with the collection surface.

Isotopic measurements were performed by secondary ion mass spectrometry (SIMS) with the Washington University Cameca IMS-3f ion microprobe. Hydrogen isotopic measurements were made by closely following previously described procedures⁸. The spatial distributions of the D/H ratios and of some elements were determined for most deuterium-rich ($\delta\text{D} \geq +1,000\text{‰}$) fragments by ion imaging. Carbon and nitrogen (measured as CN^-) isotopes were measured as negative secondary ions at a mass resolving power of $\sim 4,500$, using standard techniques¹¹.

As shown in Fig. 1, deuterium enrichments in cluster IDPs are much larger, more common and more highly variable than those in previously measured individual IDPs. Indeed, several cluster IDPs exhibit higher D/H ratios than any previously measured Solar System material, reaching 50 times the terrestrial value in the cluster IDP Dragonfly. This exceeds the highest D/H ratio measured in any meteorite¹² by nearly an order of magnitude. Dragonfly also has the largest range of D/H values among its fragments, spanning two orders of magnitude, from 9×10^{-5} to $\sim 8 \times 10^{-3}$. Most other cluster IDPs also show large variations in their D/H ratios among different fragments. An earlier ion imaging study⁹ reported substantial hydrogen isotopic variations within a fragment of the cluster IDP Butterfly. Three cluster IDP fragments imaged here exhibit localized deuterium concentrations, two of which (Roadrunner D5 and Taz D12) appear similar to the Butterfly hotspot,

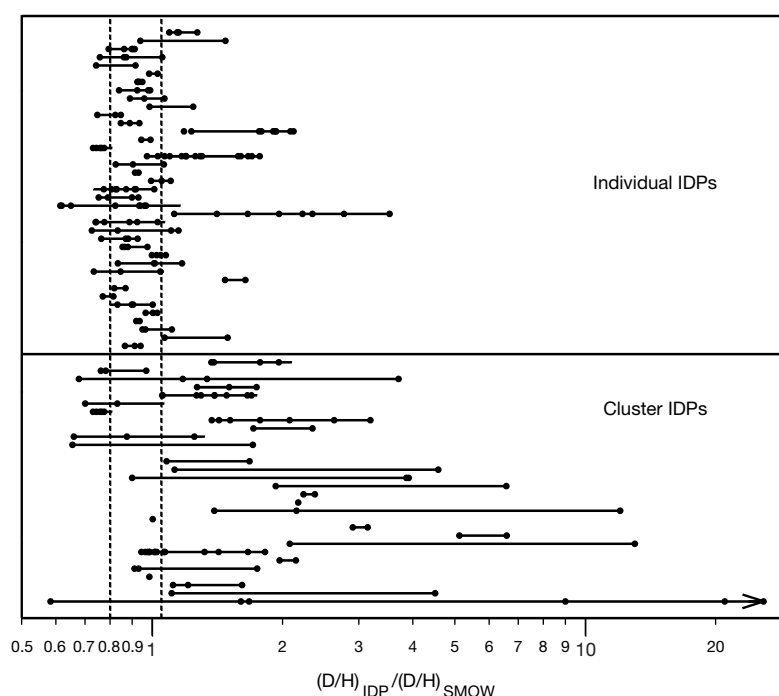


Figure 1 H isotopic ratios of IDPs. Comparison of D/H ratios measured in 40 individual IDPs⁸ and 28 cluster IDPs (ref. 8; this work). Multiple fragments of each individual and cluster IDP were measured. Each line represents the observed spread of D/H ratios among fragments of a given IDP, with the values of the fragments shown as points on the lines. The range of D/H ratios observed among terrestrial rocks is denoted by the vertical dashed lines. Both individual and cluster IDPs exhibit excess and depletions of deuterium relative to terrestrial materials although these anomalies are far more common among cluster

IDPs. Note that several fragments of each individual and cluster IDP were measured, and thus the relative abundances of D-rich fragments among the two groups is not due to a statistical sampling effect. The D/H ratios observed in cluster IDP Dragonfly (bottom line) range from the lowest to the highest values observed among extraterrestrial materials. The arrow on the Dragonfly data indicates that the maximum D/H ratio is at least 50 times the standard mean ocean water value on the basis of ion imaging.

with the D/H ratio reaching 3–4 times the value of the surrounding material (Fig. 2).

Cluster IDPs are also generally more ^{15}N -rich (mean $\delta^{15}\text{N} \approx +140\text{‰}$) than individual IDPs¹⁰ ($\sim +50\text{‰}$), including the highest ^{15}N excess so far observed in an IDP (+480‰; Dragonfly β). In one notable exception, cluster IDP Porky is substantially depleted in ^{15}N (−93‰). The nature of the isotopically light nitrogen phase is unknown, although, unlike the common ^{15}N excesses, this anomaly probably has a nucleosynthetic origin. As is the case with hydrogen, there is often a substantial range in nitrogen isotopic composition among the fragments of a given cluster.

Both individual¹⁰ and cluster IDPs have isotopically normal carbon. The mean $\delta^{13}\text{C}$ values of cluster and individual IDPs are

distinct (−45‰ and −30‰, respectively) but there is significant overlap between the groups.

It is surprising that cluster and individual IDPs are so isotopically distinct, given their many chemical and mineralogical similarities. Although it is possible that this reflects a difference in sources, the D- and ^{15}N -rich phases may simply have been lost from some particles during atmospheric entry heating. Chemical characterization of the anomalous phases would help to resolve this issue.

The extraordinarily large and highly variable hydrogen and nitrogen isotopic anomalies observed among cluster IDPs establish them as the most isotopically primitive class of Solar System materials identified to date. Although deuterium excesses are commonly observed in several types of primitive meteorites, the

Table 1 Hydrogen, carbon and nitrogen isotopic compositions of cluster IDPs

Cluster nickname	Particle ID	Cluster no.	$\delta\text{D} \pm 1\sigma$ (‰)	$\delta^{13}\text{C} \pm 1\sigma$ (‰)	$\delta^{15}\text{N} \pm 1\sigma$ (‰)
Dragonfly	L2005 A2a	31	600 $\pm$ 76	−4 $\pm$ 23	480 $\pm$ 40
	L2005 A2b	31	680 $\pm$ 44		
	L2005 F	31	20,000 $\pm$ 1,400		
	L2005 4	31	24,800 $\pm$ 1,500		
	L2005 3	31	8,000 $\pm$ 500		
	L2005 A3	31	−420 $\pm$ 24		
Speedy	L2006 A9	14	3,500 $\pm$ 280	−1 $\pm$ 20	400 $\pm$ 41
	L2006 A10	14	110 $\pm$ 46		
Penelope	L2006 A1a	4	210 $\pm$ 96	−49 $\pm$ 15	340 $\pm$ 78
	L2006 A2a	4	120 $\pm$ 29		
Gossamer	L2006 A2b	4	610 $\pm$ 77	−31 $\pm$ 13	260 $\pm$ 24
	L2006 A8	10	−14 $\pm$ 51		
	L2006 A4a	6	750 $\pm$ 52		
Claude	L2006 A3	6	−69 $\pm$ 49	−1 $\pm$ 20	400 $\pm$ 41
	L2006 A6	9	−89 $\pm$ 85		
Tweety	L2008 B2a	4	1,100 $\pm$ 92	−45 $\pm$ 11	240 $\pm$ 15
	L2008 B1	4	970 $\pm$ 56		
Elmer	L2008 310x	5	71 $\pm$ 75	−4 $\pm$ 7	89 $\pm$ 27
	L2008 310d	5	66 $\pm$ 75		
	L2008 310e	5	17 $\pm$ 79	−11 $\pm$ 7	41 $\pm$ 29
	L2008 310f	5	−14 $\pm$ 72		
	L2008 310g	5	−35 $\pm$ 6	33 $\pm$ 9	63 $\pm$ 37
	L2008 310L	5	−55 $\pm$ 18		
	L2008 310t	5		−7 $\pm$ 6	73 $\pm$ 23
	L2008 313	5	322 $\pm$ 67		
	L2008 19a	5	664 $\pm$ 84	−23 $\pm$ 27	139 $\pm$ 43
	L2008 110	5	822 $\pm$ 91		
	L2008 37a	5	−24 $\pm$ 17	−17 $\pm$ 12	130 $\pm$ 36
	L2008 37b	5	25 $\pm$ 22		
	L2008 37c	5	−14 $\pm$ 36	−16 $\pm$ 26	260 $\pm$ 34
	L2008 37d	5	12 $\pm$ 29		
Wiley	L2009 D9	10	12,000 $\pm$ 650	−18 $\pm$ 10	180 $\pm$ 18
	L2009 D10	10	1,100 $\pm$ 130		
Taz	L2009 D11	13	5,600 $\pm$ 770	−52 $\pm$ 17	240 $\pm$ 31
	L2009 D12	13	4,100 $\pm$ 570		
Marvin	L2009 F2	13	111 $\pm$ 18	−42 $\pm$ 11	190 $\pm$ 15
	L2009 D1	3	2,100 $\pm$ 120		
	L2009 D2a	3	1,900 $\pm$ 80	−120 $\pm$ 42	105 $\pm$ 35
Pepe	L2009 D2b	3			
	L2009 D4	4	3 $\pm$ 50	−52 $\pm$ 11	320 $\pm$ 33
Roadrunner	L2009 D5	7	11,000 $\pm$ 970		
	L2009 D6	7	1,200 $\pm$ 120	−34 $\pm$ 11	140 $\pm$ 13
Foghorn	L2009 E2	7	390 $\pm$ 120		
	L2009 D7	8	1,200 $\pm$ 200	−43 $\pm$ 17	140 $\pm$ 23
Yosemite	L2011 A1	6	1,200 $\pm$ 130		
	L2011 A2	6	1,400 $\pm$ 150	−1 $\pm$ 20	150 $\pm$ 27
Porky	L2011 A3	7	5,600 $\pm$ 220		
	L2011 A4	7	930 $\pm$ 89	−38 $\pm$ 15	180 $\pm$ 12
Daffy	L2011 A5	11	2,900 $\pm$ 220		
	L2011 A6	11	2,900 $\pm$ 650	−14 $\pm$ 87	100 $\pm$ 32
Petunia	L2011 A7	15	130 $\pm$ 130		
	L2011 A8	15	3,600 $\pm$ 370	−53 $\pm$ 2	−93 $\pm$ 4
Bugs	L2011 A9	22	80 $\pm$ 80		
	L2011 A10	22	680 $\pm$ 61	−55 $\pm$ 60	−10 $\pm$ 48
Leghorn	L2011 2	5	710 $\pm$ 72		
	L2011 4	5	−350 $\pm$ 25	−27 $\pm$ 14	94 $\pm$ 17
				33 $\pm$ 26	290 $\pm$ 90
				−76 $\pm$ 23	17 $\pm$ 35
				−67 $\pm$ 21	47 $\pm$ 24

Delta (δ) values represent the deviation of the measured isotopic ratio from the value of a standard material. For example, delta values for H are referenced to the D/H ratio of standard mean ocean water (SMOW) in parts per thousand (‰):

$$\delta\text{D} = \left[\frac{(\text{D/H})_{\text{sample}}}{(\text{D/H})_{\text{SMOW}}} - 1 \right] \times 1,000$$

The cluster number and individual fragment designation assigned by the NASA JSC curator are given together with the nicknames assigned for purposes of discussion. The particles included in this survey were selected only on the basis of having sufficient material (>2.5–10 μm fragments), and with at least one fragment of the cluster classified as ‘chondritic’¹³ (within a factor of three of CI chondrite elemental abundances). Hydrogen isotopes were measured as negative secondary ions, produced by a 15 keV Cs⁺ primary ion beam at low mass resolving power. Each cluster fragment was immersed in a bath of hexane for 60 s to remove silicone oil from the collector. The fragments were then individually dry transferred to a sputter cleaned gold substrate. Elemental abundances were determined by scanning electron microscopy/energy dispersive X-ray analysis.

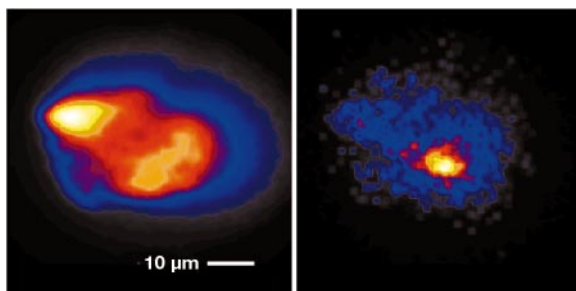


Figure 2 Hydrogen isotopic images of a deuterium-rich IDP. Secondary ion images showing the spatial distributions of H (left) and D (right) within IDP Roadrunner (D5), whose bulk D/H ratio is ~ 12 times the terrestrial value. The D (but not the H) is clearly concentrated in a small region ($< 3 \mu\text{m}$) of this IDP. The size of this D-rich 'hotspot' is similar to the spatial resolution of the ion imaging system ($1\text{--}2 \mu\text{m}$), and thus the estimated D/H ratios of this region ($3\text{--}4$ times the bulk D/H) is a lower limit. No clear correlations were observed between the distribution of D/H and other elements (for example, C, O, CN, Si, S) also imaged. Images were recorded by a Photometrics CCD camera coupled to the microchannel plate/fluorescent screen of the ion microprobe. Images were obtained at low mass resolving power. A $1\text{--}3 \text{ nA Cs}^+$ primary ion beam was defocused to $\sim 100 \mu\text{m}$ on the particles, whose dimensions ranged from $10 \mu\text{m}$ to $40 \mu\text{m}$. In most cases, a contrast aperture was used to improve the spatial resolution of the images. Hydrogen isotopic imaging runs consisted of an alternating sequence of 100-ms H^+ images and 30-s D^+ images.

D/H ratios are generally much lower than those observed among cluster IDPs². The Renazzo meteorite has yielded the highest directly measured D/H ratio¹³ ($\sim 1.4 \times 10^{-3}$), but this value is still significantly lower than those reached by cluster IDPs. Matrix material in Renazzo also exhibits some spatial variation in D/H ratio¹⁴, but among cluster IDPs the isotopic variations are both larger and at a finer spatial scale. Only two molecules have been measured in comets so far (H_2O and HCN), and although both D/H ratios are significantly enhanced ($\sim 3 \times 10^{-4}$, 2×10^{-3} respectively^{15–17}) their values are eclipsed by the D/H ratio observed in the IDP Dragonfly. Given the limited isotopic measurements of both comets and IDPs, the ranges of D/H ratios observed should be considered lower limits.

Enrichments in ^{15}N are also commonly observed in primitive meteorites. Organic compounds in primitive meteorites are ^{15}N -rich, with $\delta^{15}\text{N}$ reaching 260‰^{4,18}. Isotopic variations among the insoluble organics have been ascribed to parent-body processing of pre-solar materials¹⁸. The main nitrogen-bearing phases in several cluster IDPs have been found to be carbonaceous as well¹⁹; however, the substantial micrometre-scale isotopic variations in cluster IDPs indicate minimal processing of the original chemical components.

As shown in Fig. 3, the D/H ratio reached by cluster IDPs fall within the range of some interstellar molecules. The relatively high densities and cosmic-ray-induced ionization of molecules in these clouds drive vigorous, rapidly evolving gas-phase and grain surface chemistry. The high D/H ratios of the molecules are due to isotopic fractionation during very low temperature chemical reactions²⁰. Chemical fractionation may also occur for other light elements (C, N, O)^{21,22}; however, because of the much smaller difference in the isotopic masses, the degree of fractionation should be much smaller and may only occur at very low temperatures. For example, the $\sim 50\%$ excess of ^{15}N observed in Dragonfly would imply a formation temperature of $\sim 20 \text{ K}$ (ref. 22). Unfortunately, observational difficulties have precluded precise N, C or O isotopic measurements of molecules in cold molecular clouds. It is important to note that only simple molecules are amenable to isotopic measurements by astronomical techniques, and thus cluster IDPs probably contain a more complete sample of interstellar materials.

Although cluster IDPs share many similarities to cometary materials (for example, isotopic compositions, high carbon and volatile element abundances), an asteroidal origin cannot be

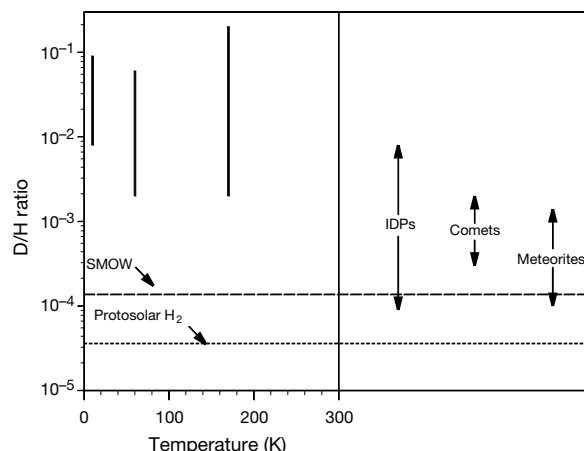


Figure 3 D/H ratios observed in interstellar molecules and in the Solar System. Left panel, D/H ratios are shown for selected molecules in TMC-1 (10 K), Orion extended ridge (50–70 K), and the Orion hot core (150–200 K). Right panel, in extraterrestrial materials the greatest range of D/H ratios and largest values are observed among IDPs, which reach those of some molecules in cold molecular clouds. Data for interstellar molecules are taken from a compilation by Millar and Hatchell²⁵. Hydrogen isotopic measurements of meteorites are discussed in ref. 2 and references therein.

excluded at present. Some individual IDPs similar to the cluster IDPs studied here (anhydrous, porous, carbon-rich) have been linked to comets on the basis of their high atmospheric entry velocities²³. Unfortunately, this criterion cannot be used for cluster IDPs owing to uncertainty in their original sizes and densities. In the future, regular sampling of the stratosphere could allow the identification of particles from specific Earth-crossing comets or asteroids²⁴.

Whatever their parent-body origins, cluster IDPs represent the most pristine material surviving from the earliest stages of Solar System formation. The components in these particles are original building blocks of the Solar System, possibly including the precursor materials of chondrules, calcium–aluminium-rich inclusions or other nebular condensates observed in primitive meteorites. Ultimately, the identification of the isotopically anomalous phases may provide a new means of studying interstellar chemical processes in the laboratory. The identification of the organic compounds present in IDPs will also help to elicit the dominant processes responsible for modifying molecular cloud compounds to those now present in primitive meteorites. □

Received 3 November 1999; accepted 10 March 2000.

1. Zinner, E. Stellar nucleosynthesis and the isotopic composition of presolar grains from primitive meteorites. *Annu. Rev. Earth Planet. Sci.* **26**, 147–188 (1998).
2. Messenger, S. & Walker, R. M. in *Astrophysical Implications of the Laboratory Study of Presolar Materials* (eds Bernatowicz, T. J. & Zinner, E.) 545–564 (AIP Conf. Proc. 402, Am. Inst. Phys, Woodbury, New York, 1998).
3. Bradley, J. P., Sandford, S. A. & Walker, R. M. in *Meteorites and the Early Solar System* (eds Kerridge, J. & Matthews, M. S.) 861–895 (Univ. Arizona Press, Tucson, 1987).
4. Sandford, S. A. The collection and analysis of extraterrestrial dust particles. *Fund. Cosmic Phys.* **12**, 1–73 (1987).
5. Flynn, G. J. et al. The volatile content of anhydrous interplanetary dust. *Meteoritics* **28**, 349–350 (1993).
6. Thomas, K. L. et al. Carbon abundance and silicate mineralogy of anhydrous interplanetary dust particles. *Geochim. Cosmochim. Acta* **57**, 1551–1556 (1993).
7. McKeegan, K. D., Walker, R. M. & Zinner, E. Ion microprobe isotopic measurements of individual interplanetary dust particles. *Geochim. Cosmochim. Acta* **49**, 1971–1987 (1985).
8. McKeegan, K. D. Ion microprobe measurements of H, C, O, Mg, and Si isotopic abundances in individual interplanetary dust particles. PhD thesis, Washington Univ., St. Louis (1987).
9. McKeegan, K. D., Swan, P., Walker, R. M., Wopenka, B. & Zinner, E. Hydrogen isotopic variations in interplanetary dust particles. *Lunar Planet. Sci.* **18**, 627–628 (1987).
10. Stadermann, F. J. Messung von isotonen- und elementen ufigkeiten in einzelnen interplanetaren staubteilchen mittels sekund rionen-massenspektrometrie. PhD thesis, Univ. Heidelberg, Germany (1991).
11. Zinner, E., Tang, M. & Anders, E. Interstellar SiC in the Murchison and Murray meteorites—isotopic composition of Ne, Xe, Si, C, and N. *Geochim. Cosmochim. Acta* **53**, 3273–3290 (1989).

12. Guan, Y., Hofmeister, A., Messenger, S. & Walker, R. M. Two types of deuterium-rich carriers in Renazzo matrix. *Lunar Planet. Sci.* **29**, abstract no. 1760 (1998).
13. Fraundorf, P., McKeegan, K. D., Sandford, S. A., Swan, P. D. & Walker, R. M. An inventory of particles from stratospheric collectors: extraterrestrial and otherwise. *J. Geophys. Res.* **87**, A403–A408 (1982).
14. Guan, Y., Messenger, S. & Walker, R. M. The spatial distribution of D-enrichments in Renazzo matrix. *Lunar Planet. Sci.* **28**, 489–490 (1997).
15. Eberhardt, P., Reber, M., Krankowski, D. & Hodges, R. R. The D/H and $^{18}\text{O}/^{16}\text{O}$ ratios in water from comet P/Halley. *Astron. Astrophys.* **302**, 301–316 (1995).
16. Bockelee-Morvan, D. *et al.* Deuterated water in comet C/1996 B2 (Hyakutake) and its implications for the origin of comets. *Icarus* **133**, 147–162 (1998).
17. Meier, R. *et al.* Deuterium in comet C/1995 O1 (Hale-Bopp): Detection of DCN. *Science* **279**, 1707–1710 (1998).
18. Alexander, C. M. O'D. *et al.* The origin of chondritic macromolecular organic matter: A C and N isotope study. *Meteor. Planet. Sci.* **33**, 603–622 (1998).
19. Keller, L. P., Messenger, S., Miller, M. & Thomas, K. L. Nitrogen speciation in a ^{15}N enriched interplanetary dust particle. *Lunar Planet. Sci.* **28**, abstract no. 1811 (1997).
20. Tielens, A. G. G. M. in *Astrophysical Implications of the Laboratory Study of Presolar Materials* (eds Bernatowicz, T. J. & Zinner, E.) 523–544 (AIP conf. Proc. 402, Am. Inst. Phys., Woodbury, New York, 1998).
21. Langer, W. D., Graedel, T. E., Frerking, M. A. & Armentrout, P. B. Carbon and oxygen isotope fractionation in dense interstellar clouds. *Astrophys. J.* **277**, 581–604 (1984).
22. Adams, N. G. & Smith, D. N-14N-15 isotope fractionation in the reaction $\text{N}_2\text{H}^+/\text{N}_2$ —interstellar significance. *Astrophys. J.* **247**, L123–L125 (1981).
23. Brownlee, D. E. *et al.* Identification of individual cometary IDPs by thermally stepped He release. *Lunar Planet. Sci.* **26**, 183–184 (1995).
24. Messenger, S. & Walker, R. M. Possible association of isotopically anomalous cluster IDPs with comet Schwassmann–Wachmann 3. *Lunar Planet. Sci.* **29**, abstract no. 1906 (1998).
25. Millar, T. J. & Hatchell, J. Chemical models of hot molecular cores. *Faraday Disc.* **109**, 15–30 (1998).

Acknowledgements

This research was supported by NASA. I thank R. Walker, E. Zinner, L. Nittler, L. Keller, J. Bradley and S. Clemett for useful discussions.

Correspondence and requests for materials should be addressed to the author.
(e-mail: dbunny@howdy.wustl.edu).

Quantum suppression of superconductivity in ultrathin nanowires

A. Bezryadin, C. N. Lau & M. Tinkham

Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA

It is of fundamental importance to establish whether there is a limit to how thin a superconducting wire can be, while retaining its superconducting character—and if there is a limit, to determine what sets it. This issue may also be of practical importance in defining the limit to miniaturization of superconducting electronic circuits. At high temperatures, the resistance of linear superconductors is caused by excitations called thermally activated phase slips^{1–4}. Quantum tunnelling of phase slips is another possible source of resistance that is still being debated^{5–8}. It has been theoretically predicted⁸ that such quantum phase slips can destroy superconductivity in very narrow wires. Here we report resistance measurements on ultrathin (≤ 10 nm) nanowires produced by coating carbon nanotubes with a superconducting Mo–Ge alloy. We find that nanowires can be superconducting or insulating depending on the ratio of their normal-state resistance (R_N) to the quantum resistance for Cooper pairs (R_Q). If $R_N < R_Q$, quantum tunnelling of phase slips is prohibited by strong damping, and so the wires stay superconducting. In contrast, we observe an insulating state for $R_N > R_Q$, which we explain in terms of proliferation of quantum phase slips and a corresponding localization of Cooper pairs.

The phenomenon of superconductivity depends on the coherence of the phase (φ) of the superconducting order parameter. For

small systems, such as Josephson junctions or ultrathin wires, φ is a quantum variable which may or may not have a definite value, corresponding to superconducting and insulating states respectively: the system becomes superconducting when its wavefunction becomes localized in the φ -space. Properties of a Josephson junction can be understood from an analogy with a quantum particle in a periodic potential⁹, whose wavefunction is a delocalized Bloch wave. Since the Josephson energy is a periodic function of φ , the Josephson junction should also be delocalized in the φ -space, and thus insulating, for arbitrary strength of the Josephson coupling. But this is not always true, as the Josephson junction is a macroscopic quantum system¹⁰ which interacts with its environment. This interaction¹¹, when linear, can be described in terms of the classical friction coefficient (η). The friction can reduce the energy bandwidth to zero and localize the particle^{12–14}. Such a quantum localization transition is called a dissipative phase transition (DPT)¹⁵. It occurs at some critical value of η independent of the strength of the periodic potential. In the case of Josephson junctions¹⁵, the dissipation is controlled by the normal conductance (G_{NJ}), and the DPT appears as a superconductor–insulator transition at $G_{NJ} = (2e)^2/h = 1/R_Q$, where e is the charge of an electron, and h is Planck's constant. The physics of superconducting wires is more complicated because φ can vary along the wire. The nature of superconductor–insulator transitions in nanowires has recently been analysed theoretically by many authors^{7,8,16}, yet there is no consensus at present.

Here we investigate experimentally the possibility of a DPT in ultrathin wires. If it exists, it could be the main mechanism which controls superconductivity in nanowires. By analogy with Josephson junctions, we may expect that the DPT should be controlled by the wire's normal-state conductance (G_N , equivalent to $1/R_N$) and should not depend explicitly on its diameter, which determines the energy barrier for phase slips. We have measured several nanowires

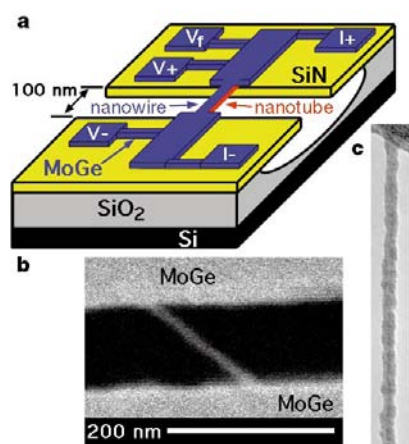


Figure 1 Fabrication and imaging of nanowire. **a**, Diagram of the sample. The Si substrate (shown black) is covered with a 0.5- μm layer of SiO_2 and a 50-nm SiN film. A 100-nm-wide slit is patterned in the SiN film using electron beam lithography and reactive ion etching. The SiO_2 layer under the slit is removed using HF to form an undercut (shown white). To make a nanowire, we first deposit nanotubes (red) and then sputter a 5-nm-thick amorphous $\text{Mo}_{79}\text{Ge}_{21}$ film and a 1.5-nm Ge protective layer. Initially the metal film covers the entire substrate including free-standing carbon nanotubes or bundles crossing the slit. Next we find a suitable single nanowire (using scanning electron microscopy, SEM) and pattern the electrodes (blue) using optical lithography and reactive ion etching. The Mo–Ge film is interrupted by the slit, and forms two electrodes connected electrically through a single nanowire. **b**, SEM image of a sample, showing a free-standing nanowire which connects two Mo–Ge electrodes. This is one of the thinnest wires found under the SEM. Its apparent width, including blurring in the SEM image, is $W \approx 10$ nm. The scale bar (white) is 200 nm. **c**, One of the thinnest wires found under a higher-resolution transmission electron microscope. The wire width is $W_{\text{TEM}} \approx 5.5 \pm 1$ nm. The scale bar (black) is 100 nm.

of diameters less than or equal to 10 nm, ranging in length L from ~ 95 nm to ~ 185 nm, much longer than the coherence length $\xi = 8$ nm. It is found that the wires are superconducting only if R_N is lower than the quantum resistance for Cooper pairs ($R_q = h/(2e)^2 \approx 6.5$ k Ω), and insulating otherwise. This agrees with the DPT interpretation: at weak dissipation, the tunnelling of phase slips is dominant in ultrathin wires and destroys superconductivity. Nevertheless, the superconductivity is recovered at $R_N < R_q$, when the dissipation, proportional to G_N , is strong enough to suppress the quantum phase slip (QPS) tunnelling.

Quantum effects are measurable only in ultrathin nanowires of size ~ 10 nm (ref. 8), which is below the resolution limit of electron beam lithography. We have developed a technique which allows fabrication of uniform nanowires considerably thinner than 10 nm (Fig. 1c). This is achieved by sputtering a superconducting alloy of amorphous $\text{Mo}_{79}\text{Ge}_{21}$ over a free-standing carbon nanotube (or bundle of tubes) which is laid down over a narrow and deep slit¹⁷ etched in the substrate (Fig. 1a). The wire width is determined by the width of the underlying bundle, which serves as a template for metal deposition. Sputtered Mo–Ge films are amorphous, have a sharp superconducting transition, and show no signs of granularity down to ~ 1 nm film thickness (ref. 18). Our nanowires are five times thicker, so they are expected to be very homogeneous. Indeed, even the narrowest wires of width $W \approx 5.5$ nm are continuous (see Fig. 1c), with a surface roughness of ~ 1 nm.

The sample resistance was determined from the slope of the current–voltage (I – V) curves measured at a frequency of 0.48 Hz by current biasing the leads $I+$ and $I-$ (Fig. 1a). A low bias amplitude (4 nA) ensured the linearity of the I – V curves. The voltage was measured with a low-noise battery-operated amplifier PAR 113 (Princeton Applied Research). Examples of resistance versus temperature curves are shown in Fig. 2a. The bottom curve shows the superconducting transition of the electrodes. The top curve gives

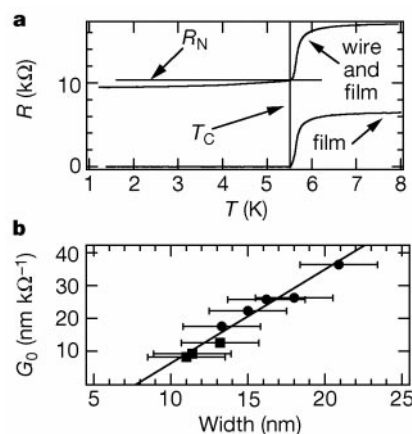


Figure 2 Normal-state properties of the nanowires. **a**, Definition of the normal-state wire resistance (R_N). The bottom curve is the resistance of the 5-nm-thick Mo–Ge film (the leads) versus temperature, measured on the contacts V_1 and V_+ (see Fig. 1a). The top curve is taken on the V_+ and V_- contacts, and shows the resistance $R(T)$ of a nanowire connected in series with a section of the leads. Below 5.5 K (vertical line) the leads are superconducting and the top curve gives the resistance of the nanowire. The normal-state resistance R_N (horizontal line) is measured immediately below the film transition where the wire is still normal. **b**, Unit-length wire conductance $G_0 = L/R_N$ versus the wire width W . The width is measured by SEM. L is the effective length of the wire. Each data point represents a different sample. The data are shown for superconducting (circles) and insulating (squares) samples. The straight line is $G_0 = C(W - W_0)$ with $C = 2.8 \times 10^{-3}$ S and $W_0 = 7.9$ nm. Note that W systematically overestimates the actual width of the conducting core of the wire owing to SEM smearing (compare Fig. 1b and c) and the presence of the Ge protective film on top of each wire. This explains why the linear fit does not extrapolate to zero.

the resistance $R(T)$ of the nanowire in series with a section of the electrodes. Below 5.5 K the electrodes are resistanceless but the wire is not, and $R(T)$ exactly equals the wire resistance. The resistance measured immediately below the film transition is taken to be the normal-state resistance R_N of the wire¹⁹ (Fig. 2a).

The homogeneity (absence of a granular structure) of the wires is established by measuring their normal-state bulk resistivity. We find that the resistivity is the same for all wires, and agrees well with the standard values for amorphous $\text{Mo}_{79}\text{Ge}_{21}$. In Fig. 2b we show a plot of the unit-length normal conductance $G_0 \equiv L/R_N$ versus the wire width W . The width is measured using a scanning electron microscope. An example is shown in Fig. 1b. All the data points in Fig. 2b, including the insulating samples, can be fitted with a straight line. The slope gives the resistivity as $\rho = d/(dG_0/dW) \approx (1.8 \pm 0.4) \mu\Omega$ m. This is in excellent agreement with the values $\rho_f \approx 1.7 \mu\Omega$ m and $\rho_b \approx 1.65 \mu\Omega$ m measured on 5-nm-thick amorphous $\text{Mo}_{79}\text{Ge}_{21}$ films¹⁸ and on bulk samples, respectively. Here we have assumed that the film thickness is constant across the wire and equals the film thickness $d = 5$ nm. Some deviations from this model geometry are possible, as the film on the supposedly circular nanotubes can have tapering edges. On the other hand, the deviations should be weak because the sputtering is far from being unidirectional. Also, the edge roughness is much smaller than the coherence length, so the superconducting effects should not be influenced. The following measurements also confirm that the wires are uniform over their length. (1) The residual resistance ratio for all

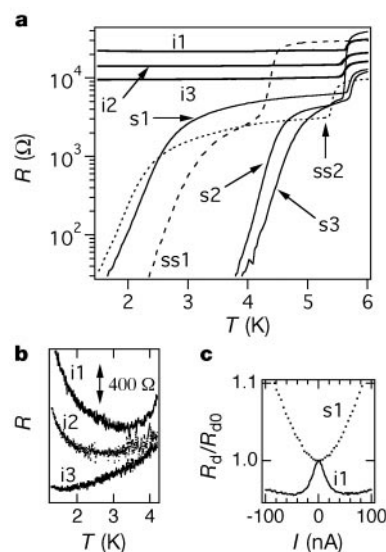


Figure 3 Transport properties of superconducting and insulating nanowires. **a**, Resistance versus temperature curves for eight different samples. The superconducting transition of the leads takes place at $T_c \approx 5.5$ K. Sample ss1 has a different T_c (≈ 4.3 K), presumably due to a different substrate treatment. Samples ss1 (dashed curve) and ss2 (short-dash curve) contain two parallel wires. All other samples have only a single wire. The samples i1, i2, i3, s1, s2, s3, ss1 and ss2 have the following parameters. Apparent widths (nm; measured under SEM) are $W = 11, 11.4, 13.2, 18, 21, 16.2, 13.3$ and 15 , respectively. The wires in each of pair ss1 and ss2 have about the same width. The normal-state resistances (k Ω) are $R_N = 22.6, 14.79, 10.29, 6.42, 4.53, 5.66, 3.09$ and 3.2 , respectively. The effective lengths (nm) are $L = 185, 135, 130, 168, 165, 146, 57$ and 72 , respectively. For the pairs of parallel wires, the quoted length is calculated as $1/L = 1/L_1 + 1/L_2$. The lengths of the ss1 wires are $L_1 = 96$ nm and $L_2 = 139$ nm. For the ss2 pair, the lengths are $L_1 \approx L_2 \approx 143$ nm. **b**, Magnification of the $R(T)$ curves i1, i2 and i3. The curves are vertically displaced for clarity. All three samples show $dR/dT < 0$ at low enough temperatures. The upturn takes place at $T = 3.2, 2.8$ and 1.6 K for the samples i1, i2 and i3, respectively. **c**, A plot of normalized differential resistance ($R_d = dV/dI$, $R_{d0} = dV/dI|_{I=0}$) versus the bias current (I) for two samples: i1 ($T = 1.2$ K) and s1 ($T = 4.2$ K).

wires is ≥ 0.85 , similar to the standard macroscopic value of ~ 0.95 . (2) A single-electron-tunnelling gate effect has not been seen (the doped Si substrate served as a gate). (3) The critical current was $\sim 1 \mu\text{A}$ (for sample s3; see below) which is similar to the value expected for a uniform wire. (4) The transition at the critical current is very sharp and shows no steps.

The results of our transport measurements on nanowires are summarized in Fig. 3a. Two qualitatively different types of behaviour are found. The samples s1, s2, s3, ss1 and ss2 appear to be superconducting as their resistance decreases approximately exponentially with cooling. The wires i1, i2 and i3 (thick curves) do not show any signatures of superconductivity (the resistance drop at $\sim 5.5 \text{ K}$ comes from the electrodes), and their resistance stays almost constant with cooling. Therefore these samples may be normal or even insulating. In fact, they should be considered insulating because their resistance increases at low enough temperatures (Fig. 3b), and their differential resistance ($R_d \equiv dV/dI$) has a pronounced maximum¹⁵ (Fig. 3c, bottom curve) at zero bias current (I), and therefore $dR_d/dI < 0$ at small I . The superconducting samples showed positive derivatives $dR/dT > 0$ and $dR_d/dI > 0$ (Fig. 3c, top curve) in all cases, including very low temperatures: one sample has been tested down to 50 mK.

The data in Fig. 3a and b show that the type of the $R(T)$ dependence is controlled by the normal resistance R_N of the wire (see also Fig. 2a). This behaviour can be explained assuming that the wires undergo a superconductor–insulator quantum phase transition at $R_N = R_q$. This conclusion is supported by the following facts. (1) All samples with $R_N < R_q$ are superconducting while all samples with $R_N > R_q$ are insulating. (2) The insulating as well as superconducting properties become stronger when the difference $|R_N - R_q|$ increases (except for the samples with two wires). For example, the temperature of the upturn in Fig. 3b increases with R_N . (3) The difference between the superconducting and insulating samples becomes more pronounced at lower temperatures, indicating that the transition is driven by quantum (not thermal) fluctuations. Note that the observed dichotomy (Fig. 3a) resembles the superconductor–insulator transition in amorphous thin films (the two-dimensional limit)²⁰ which occurs when their resistance per square $R_{\square} = \rho/d$ reaches $R_q \approx 6.5 \text{ k}\Omega$. Despite apparent similarities, we think that our transition is different, as our 5-nm-thick Mo–Ge film (which forms the leads and the wire) has a much lower resistance per square $R_{\square} \approx 350 \Omega \ll R_q$.

Two dashed curves in Fig. 3a show data for samples with two parallel wires. As expected, they have the lowest normal-state resistance. Nevertheless, their low-temperature tails are similar to single-wire samples with about twice higher normal resistance. This shows that the dissipative events (presumably the phase slips) are localized inside each wire, and do not depend on the presence of another wire $\sim 0.5 \mu\text{m}$ away.

The important role of the normal-state conductance $G_N \approx 1/R_N$, evident from our measurements, suggests that the observed superconductor–insulator transition is a DPT. The amount of dissipation is controlled by the normal conductance G_N because each phase slip generates a finite voltage pulse (V_p) on the wire. The dissipated power is simply $V_p^2 G_N$, that is, proportional to the normal conductance. Each phase slip transforms the kinetic energy of Cooper pairs into thermal energy and causes heating. At zero temperature and zero bias current, virtual phase slips need to be considered. Absence of real phase slips ensures that no real dissipation occurs when the system is in its ground state. In this case the damping, which is still proportional to the normal conductance, is also virtual.

As we have already discussed, a quantum system becomes superconducting when the damping is strong enough to suppress tunnelling in the φ -space and to localize the phase. In contrast to the case of a Josephson junction (which is geometrically a zero-dimensional system), the phase fluctuations in nanowire (which is

one-dimensional) can vary in length from ξ up to the length of the wire (L). The ‘friction’ acting on a QPS depends on the length scale over which the phase change occurs. The longer the QPS, the weaker the dissipation because the conductance is proportional to $1/L$. The tunnelling of larger QPSs can therefore be dominant, and the mean field theory, which considers only phase slips of size ξ (which are energetically the most favourable) may not be sufficient to analyse the DPT. To accurately describe all sizes of QPS, it is necessary to use the renormalization group theory of Zaikin *et al.*⁸ and E. Demler (personal communication). Qualitatively speaking, the QPS can be suppressed at all scales, up to L , only if the normal conductance of the whole wire exceeds the conductance quantum. Therefore, independently of the wire length, the DPT should take place at $G_N = (2e)^2/h$. Our measurements on relatively long wires ($L \approx 20\xi$) appear to be in good agreement with the above conclusions, confirming that we observe a DPT. Although it may seem surprising that a DPT in nanowires occurs at the same resistance ($R_N = R_q$) as in Josephson junctions, this could be a natural consequence of the critical-point divergence of the length scale of fluctuations. When the spatial dimensions of the quantum fluctuations become larger than the length of the wire, the system (nanowire) becomes effectively zero-dimensional and therefore equivalent to Josephson junction. This shows a similarity between quantum phase transitions observed in mesoscopic systems and thermodynamic phase transitions which occur in bulk systems.

The negative derivatives $dR/dT < 0$ and $dR_d/dI < 0$, measured on the insulating wires, can also be understood in terms of QPSs. The insulating state is delocalized in the φ -space and localized in the conjugate, charge space, meaning localization of the Cooper pairs. External perturbations such as thermal fluctuations or a nonzero bias voltage weaken the charge-localization effect, leading to the observed reduction in the differential resistance. The shape of the dR_d/dI versus I dependence measured for the insulating samples (Fig. 3c) is very similar to measurements¹⁵ on Josephson junctions which undergo a DPT. This similarity gives an additional confirmation that the transition which we find near $G_N = (2e)^2/h$ is a DPT.

We plan to investigate the DPT in longer wires, and to compare it to the renormalization group theory of Zaikin *et al.*⁸ which predicts that the wire diameter becomes important when the length exceeds $\sim \hbar c_{\text{MS}}/k_B T \approx 10 \mu\text{m}$. Here c_{MS} is the velocity of the Mooij–Schön mode²¹, which is an electromagnetic wave that propagates along the wire. □

Received 6 January; accepted 25 February 2000.

1. Langer, J. S. & Ambegaokar, V. Intrinsic resistive transition in narrow superconducting channels. *Phys. Rev.* **164**, 498–510 (1967).
2. McCumber, D. E. & Halperin, B. I. Time scale of intrinsic resistive fluctuations in thin superconducting wires. *Phys. Rev. B* **1**, 1054–1070 (1970).
3. Tinkham, M. *Introduction to Superconductivity* 2nd edn, 288 (McGraw Hill, New York, 1996).
4. Ivlev, B. I. & Kopnin, N. B. Electric currents and resistive states in thin superconductors. *Adv. Phys.* **33**, 47–114 (1984).
5. Van Run, A. J., Romijn, J. & Mooij, J. E. Superconducting phase coherence in very weak aluminium strips. *Jpn J. Appl. Phys.* **26**, 1765–1766 (1987).
6. Giordano, N. Evidence for macroscopic quantum tunneling in one-dimensional superconductors. *Phys. Rev. Lett.* **61**, 2137–2140 (1988).
7. Duan, J.-M. Quantum decay of one-dimensional supercurrent: role of electromagnetic field. *Phys. Rev. Lett.* **74**, 5128–5132 (1995).
8. Zaikin, A. D., Golubev, D. S., van Otterlo, A. & Zimanyi, G. T. Quantum phase slips and transport in ultrathin superconducting wires. *Phys. Rev. Lett.* **78**, 1552–1555 (1997).
9. Likharev, K. K. & Zorin, A. B. Theory of the Block-wave oscillations in small Josephson junctions. *J. Low Temp. Phys.* **59**, 347–382 (1985).
10. Averin, D. V. Solid-state qubits under control. *Nature* **398**, 748–749 (1999).
11. Caldeira, A. D. & Leggett, A. J. Influence of dissipation on quantum tunneling in macroscopic systems. *Phys. Rev. Lett.* **46**, 211–214 (1981).
12. Schön, G. & Zaikin, A. D. Quantum coherent effects, phase transitions, and the dissipative dynamics of ultra small tunnel junctions. *Phys. Rep.* **198**, 237–412 (1990).
13. Bulgadev, S. A. Phase diagram of a dissipative quantum system. *JETP Lett.* **39**, 315–319 (1984).
14. Schmid, A. Diffusion and localization in a dissipative quantum system. *Phys. Rev. Lett.* **51**, 1506–1509 (1983).
15. Penttilä, J. S., Hakonen, P. J., Paalanen, M. A. & Sonin, E. B. “Superconductor–insulator transition” in a single Josephson junction. *Phys. Rev. Lett.* **82**, 1004–1007 (1999).
16. Zaikin, A. D., Golubev, D. S., van Otterlo, A. & Zimanyi, G. T. Quantum fluctuations and dissipation in thin superconducting wires. *Usp. Fiz. Nauk* **168**, 244–248 (1998).

is not the case for the silicon nitride structural layer employed here; however, this material does provide a homogeneous effective medium for the long-wavelength phonons of relevance to these experiments. Recent measurements have shown that phonon propagation in this material is ballistic at low temperatures^{20,21}.

Motivated by idealized scalar-wave calculations for an isotropic elastic solid of comparable geometry and dimensions¹¹, we have patterned the width of the waveguides to follow $\cosh^2(x/\lambda)$, where x is the longitudinal coordinate and $\lambda = 1.0 \mu\text{m}$. The calculations, valid for longitudinal wave propagation through waveguides of this shape, indicate that quantized thermal conductance may be observable over an appreciable temperature range; this range is bounded at low temperatures by loss of adiabatic coupling to the reservoirs and, at higher temperatures, by the onset of thermal population of higher-energy (massive) modes.

Integration of carefully designed transducers allows thermal conductance measurements on the nanoscale phonon waveguides. We realize these by patterning (from a 25-nm-thick Au film) two thin-film resistors on top of the phonon cavity. As described below, these serve as a local heater and thermometer for the measurements;

a second thermometer measures substrate temperature. Electrical connection to these transducers is provided by thin-film Nb leads (thickness 25 nm) patterned above the waveguides; these extend from the cavity to wirebond pads anchored to the main part of the chip. By virtue of their superconductivity, these leads do not provide a parasitic thermalization path to the reservoirs.

In principle, the measurement technique is simple—we first raise the temperature of the phonon cavity by Joule-heating its integrated Au resistor. We then employ the second transducer to measure the induced change in temperature. The ratio of heat current flowing through the waveguides to the induced rise in cavity temperature is the phonon thermal conductance of the waveguides. In practice, however, this measurement is quite challenging—it is necessary to obtain a sensitive measurement of phonon temperature within a suspended nanostructure that has very small heat capacity. The measurement process itself must involve minimal power dissipation, because exceedingly weak coupling exists between the sensor electrons and the cavity phonons at very low temperatures ($10^{-13} \text{ W K}^{-1}$ at 100 mK for $0.1 \mu\text{m}^3$ of Au)²². As depicted in Fig. 2, we employ d.c. SQUID-based noise thermometry^{23,24} to measure the

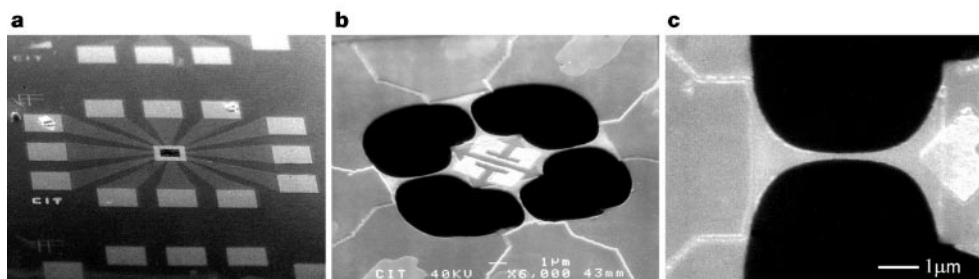


Figure 1 Suspended mesoscopic device. A series of progressive magnifications are shown. **a**, Overall view of the $\sim 1.0 \times 0.8 \text{ mm}$ device, showing 12 wirebond pads that converge via thin-film niobium leads into the centre of the device. This central region is a 60-nm-thick silicon nitride membrane, which appears dark in the electron micrograph. **b**, View of the suspended device, which consists of a $4 \times 4 \mu\text{m}$ 'phonon cavity' (centre) patterned from the membrane. In this view, the bright 'c' shaped objects on the device

are thin-film gold transducers, whereas in the dark regions the membrane has been completely removed. The transducers are connected to thin-film niobium leads that run atop the 'phonon waveguides'; these leads ultimately terminate at the wirebond pads. **c**, Close-up of one of the catenoidal waveguides, displaying the narrowest region which necks down to $<200\text{-nm}$ width.

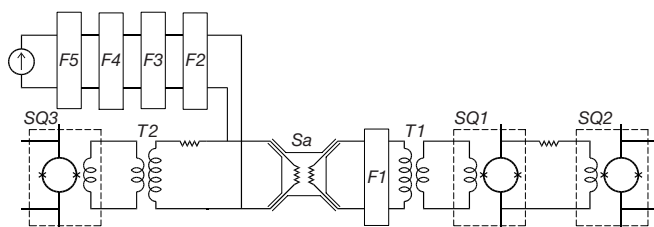


Figure 2 Simplified apparatus diagram. The suspended sample (Sa), cooled on the dilution refrigerator's mixing chamber (MC), is represented by two ($\sim 20 \Omega$) resistors in the centre of the figure. Its rightmost (sensor) transducer is coupled to a d.c. SQUID-based noise thermometry circuit. Two thin-film d.c. SQUIDs (SQ1, SQ2) are cooled on the MC to enable ultralow noise performance: SQ1 is run open-loop (with continuous gain measurement carried out at 900 Hz, that is, just outside the noise measurement band) while its readout, SQ2, is operated in flux-locked mode. The metallic powder filter (F1) on the MC shields the sample from Josephson radiation emanating from the SQUIDs. The sample's leftmost (heater) transducer is connected to a heavily filtered (F2–F5) current biasing circuit used to Joule-heat the phonon cavity. A SQUID voltmeter circuit (SQ3) on the MC allows resistance measurement *in situ*, by sensing the voltage drop across the current biased heater. The heavily shielded superconducting transformers (T1, T2) provide a current gain of ~ 18 . The cold sample is isolated from heat conduction from the higher-temperature environment through filters F2 and F3 (cooled on the MC) and F4 and F5 (at 700 mK). F2 and F4 are metallic powder filters, while F3 and F5 are 10-pole RC filters.

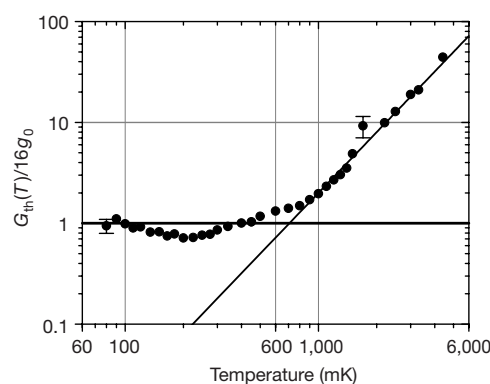


Figure 3 Thermal conductance data. We normalize the measured thermal conductance by the expected low-temperature value for 16 occupied modes, $16 g_0$. Measurement error is approximately the size of the data points, except where indicated. For temperatures above $T_{co} \approx 0.8 \text{ K}$, we observe a cubic power-law behaviour consistent with a mean free path of $\sim 0.9 \mu\text{m}$. For temperatures below T_{co} , we observe a saturation in G_{th} at a value near the expected quantum of thermal conductance. Consistent with expectations, to within experimental uncertainty, G_{th} never exceeds $16 g_0$ once it finally drops below that value. The (implicit) linear temperature dependence observed for $T < T_{co}$ is the hallmark of one-dimensional thermal transport. The approximate saturation at the maximum expected magnitude, $G_{th} \approx 16 g_0$, indicates that both the ballistic and adiabatic conditions have been satisfied in these experiments.

Acknowledgements

We thank M. C. Cross, R. Lifshitz, G. Kirczenow, M. Blencowe, N. Wingreen and P. Burke for discussions, suggestions and insights, and N. Bruckner for assistance with silicon nitride growth. We thank M. B. Ketchen and members of the IBM Yorktown superconductivity group for advice, assistance and the d.c. SQUID devices employed in our cryogenic electronics. This work was supported by DARPA MTO/MEMS and NSF/DMR.

Correspondence and requests for materials should be addressed to M.L.R. (e-mail: roukes@caltech.edu).

Coulomb-blockade transport in single-crystal organic thin-film transistors

W. A. Schoonveld^{*,†}, J. Wildeman^{†,‡}, D. Fichou[§], P. A. Bobbert^{||}, B. J. van Wees^{*,†} & T. M. Klapwijk[¶]

^{*} Department of Applied Physics, [†] Materials Science Center, [‡] Department of Polymer Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

[§] Laboratoire des Matériaux Moléculaires, CNRS, 2 rue Henry-Dunant, 94320 Thiais, France

^{||} Physics Department, Eindhoven University of Technology, PO Box 513, NL-5600 MB Eindhoven, The Netherlands

[¶] Delft University of Technology, Department of Applied Physics, Nanophysics and Nanotechnology Section, Lorentzweg 1, 2628 CJ Delft, The Netherlands

Coulomb-blockade transport—whereby the Coulomb interaction between electrons can prohibit their transport around a circuit—occurs in systems in which both the tunnel resistance, R_T , between neighbouring sites is large ($\gg h/e^2$) and the charging energy, E_C ($E_C = e^2/2C$, where C is the capacitance of the site), of an excess electron on a site is large compared to kT . (Here e is the charge of an electron, k is Boltzmann's constant, and h is Planck's constant.) The nature of the individual sites—metallic, superconducting, semiconducting or quantum dot—is to first order irrelevant for this phenomenon to be observed¹. Coulomb blockade has also been observed in two-dimensional arrays of normal-metal tunnel junctions², but the relatively large capacitances of these micrometre-sized metal islands results in a small charging energy, and so the effect can be seen only at extremely low temperatures. Here we demonstrate that organic thin-film transistors based on highly ordered molecular materials can, to first order, also be considered as an array of sites separated by tunnel resistances. And as a result of the sub-nanometre sizes of the sites (the individual molecules), and hence their small capacitances, the charging energy dominates at room temperature. Conductivity measurements as a function of both gate bias and temperature reveal the presence of thermally activated transport, consistent with the conventional model of Coulomb blockade.

The experimental transport data available in the literature^{3–8}, obtained with highly ordered polycrystalline thin-film transistors (TFTs) based on α -6T and pentacene (see Fig. 1a), have so far been interpreted in terms of hopping between localized trap states, polaron hopping (both correlated and uncorrelated) or in terms of the Holstein model. However, large variations in the experimental data—on even nominally identical samples—limit the understanding of the intrinsic charge transport mechanisms of these systems.

Our sample layout is shown in Fig. 1a. The substrate is a highly doped silicon wafer, acting as a gate electrode, which is thermally oxidized in a dry atmosphere. The gold source and drain electrodes are lithographically defined, with gap diameters ranging from 2 to

20 μm . A second isolation layer is deposited to isolate the transistor electrically from neighbouring devices. The organic materials used were pentacene (Aldrich), quaterthiophene (α -4T; synthesized by Syncom BV) and sexithiophene (α -6T; also from Syncom BV). These materials were purified, and deposited in the last fabrication step by thermal evaporation in a high-vacuum environment (1×10^{-7} mbar). Large individual crystallites within a polycrystalline thin film are obtained by optimization of the substrate temperature and evaporation rate^{9,10}. With optimal settings the individual crystallites can grow up to 40 μm in diameter, which is large enough to fill one gap completely, resulting in a single-crystal TFT¹¹. The identification, orientation and thickness of the single crystals can be obtained by an optical polarization microscopy technique⁹.

Room-temperature current–voltage (I – V) characteristics of a crystal pentacene TFT with a thin-film phase¹⁰, measured with increasing and decreasing drain voltage, are shown in Fig. 2a. All measurements are performed under high-vacuum conditions, and as a result show little or no hysteresis in the drain voltage sweep. In contrast to what is generally reported in the literature, we typically observe non-ohmic behaviour of the gold source and drain contacts to the organic active layer¹¹, as is demonstrated at low drain bias in Fig. 2. The conductivity is obtained in the linear transport regime as a function of gate bias¹², and corrected for the non-ohmic contacts. A typical result is shown in Fig. 2b. The non-ohmic effects of the contacts were compensated for by measuring several sweeps of gate voltage with increasing source–drain bias ($V_d = -4, -5, -6$ and -7 V). The channel conductance at a set gate voltage is then obtained from the slope of the channel current with the applied

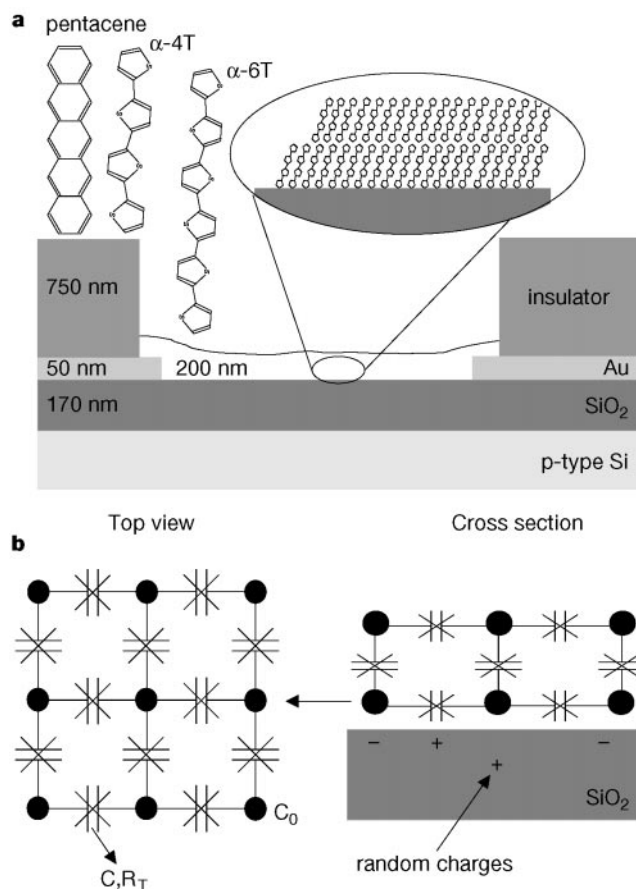


Figure 1 Schematic sample layout. **a**, Single-crystal TFT sample structure for α -4T, α -6T and pentacene materials. **b**, A schematic view of two molecular layers on top of the SiO_2 gate-dielectric with random charges incorporated. In the in-plane direction, the layers are represented by a two-dimensional array of molecules separated by a tunnel junction resistance R_T with nearest-neighbour capacitance C . The self-capacitance molecule is given by C_0 .

drain voltage. This procedure is applied to all measured gate voltages, enabling the channel conductance to be obtained as a function of gate voltage.

The resulting charge transport data of the pentacene TFT are shown in Fig. 3. The channel conductivity is clearly thermally activated, with a gate-voltage-dependent activation energy (Fig. 3a and c). Similar behaviour of the activation energy has been reported¹³ for organic TFTs with an amorphous active layer. In this case the transport has been described by a variable-range hopping process between localized trap states.

We found that the extrapolated conductivity σ_0 at high temperatures shows a linear dependence on gate voltage (Fig. 3b) and vanishes at a threshold voltage of $V_t = +9.6$ V. This threshold voltage marks the onset for p-type conduction in pentacene TFTs, and is a crucial value needed to calculate the actual charge carrier density n_s in the active layer. Note that this value contrasts sharply with the 'threshold voltage', which might be extracted from the quasi-linear regime of the channel conductance with gate voltage (Fig. 2b). The threshold voltage obtained in this fashion is in our view not relevant to the calculation of the effective charge carrier density, as no real onset for conduction can be discriminated.

The transport data are interpreted in terms of the Coulomb-blockade model in the orthodox regime, in which the tunnel resistance between neighbouring sites is larger than the quantum resistance ($R_T \gg R_Q = h/e^2 \approx 25$ k Ω)¹. In a first-order approximation, we will assume a metallic-like behaviour of the molecular islands. Note however that, in contrast to metallic islands, the level spacing δE in molecules is generally larger than kT . With this relatively large level spacing the bandwidth Γ , which is given by the product of the transmission probability and the energy level spacing¹⁴ $\Gamma \equiv |t|^2 \delta E$, could become large enough to break down the regime of the orthodox Coulomb blockade.

Transport in the orthodox regime is described by tunnelling of localized charges between nearest-neighbour islands, driven by the energy difference between the initial and final energy states. The charging energy of a site can be approximated by:

$$E_C = \frac{(ne - Q_{\text{ind}})^2}{2C} \quad (1)$$

where Q_{ind} is the induced charge on a molecule, and n the number of charge carriers on the molecule. The induced charge Q_{ind} can have

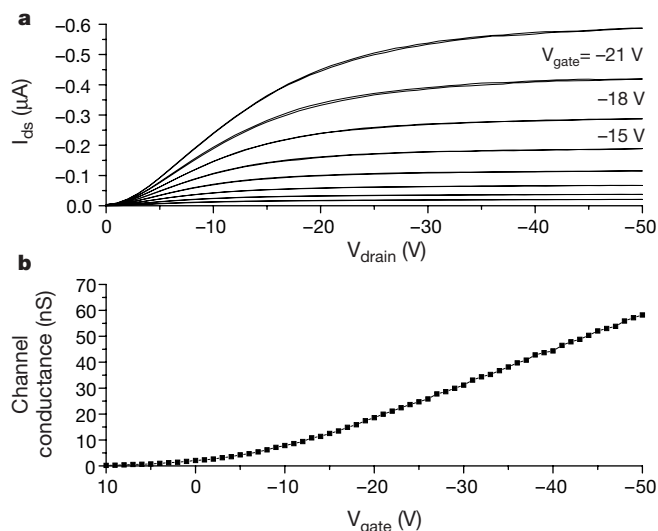


Figure 2 Room-temperature current–voltage characteristics of a single-crystal pentacene TFT with gap length of 8 μm and width of 16 μm . **a**, Standard TFT characteristics, showing the linear and saturation regimes. Note the gradual onset at low drain bias which is due to non-ohmic contacts. **b**, The channel conductance as a function of gate bias. The data have been corrected for the non-ohmic contacts.

non-integer values because of the random offset charges present in the SiO₂ dielectric layer (Fig. 1b). Alternatively, positional disorder of the molecules can give rise to different nearest-neighbour capacitances, which also affects the charging energy (equation (1)). The maximum energy barrier between neighbouring sites can be estimated by the charging energy at the degeneracy point at which the induced charge on the dot is equal to a half-filling $Q_{\text{ind}} = e(n \pm 1/2)$ (equation (1)).

It follows that the observed energy barrier is a product of the charging energy and the induced random offset charges. The charging energy of a molecule is of the order of $E_C = e^2/2C \approx 1$ eV and the typically observed activation energies are roughly 0.1 eV, from which we estimate the maximum induced charge on a molecule to be about 10% of an electron charge. The transport is then probably dominated by activated hops between nearest-neighbour sites, with an energy barrier dependent on the charge carrier density. We suggest that the conductivity can be described by:

$$\sigma(n_s, T) = \sigma_0 \exp\left(-\frac{E_{\text{act}}(n_s) + E_p}{kT}\right) \quad (2)$$

with $\sigma_0 = n_s e \mu_0$ the high-temperature conductivity, and μ_0 the intrinsic mobility determined by $\mu_0 = (R_T n_{\text{sites}} e)^{-1}$ with R_T an energy independent tunnel resistance. (The number of available sites n_{sites} is obtained from single crystal X-ray diffraction data^{15–17}

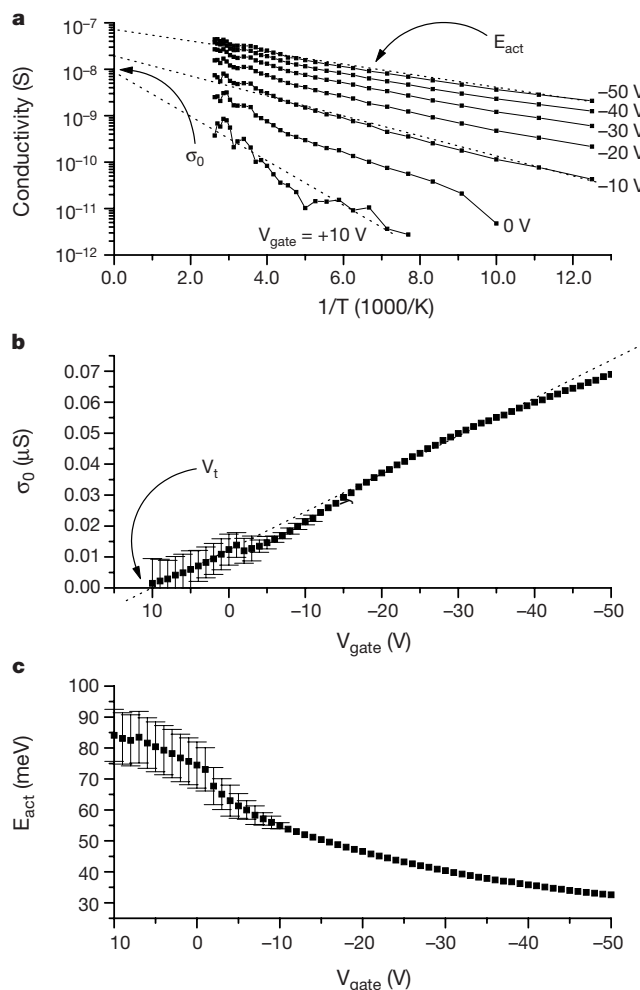


Figure 3 Charge transport data of the pentacene TFT shown in Fig. 2. **a**, The conductivity shows thermally activated behaviour over the entire gate voltage range. **b**, The extrapolated conductivity σ_0 at $T \rightarrow \infty$ increases linearly with gate voltage. The threshold voltage V_t marks the onset of p-type conduction for pentacene TFTs. **c**, The resulting activation energy as a function of gate voltage. The error bars are representative of the I – V measurement accuracy.

and is given in Table 1), n_s the charge carrier density, E_{act} the activation energy and E_p a possible polaron binding energy term. The orthodox Coulomb-blockade description assumes that the intermediate microscopic processes, such as the details of the tunnelling and relaxation processes once the charge has tunnelled, can in first order be neglected (this requires the polaron relaxation time, Δt , to be $< \hbar/E_p$). Note that this contrasts with the polaron hopping theories^{18,19}, in which the microscopic details of the hopping motion of the charges are relevant for the charge transport mechanism. The activation energy is dependent on the charge carrier density and the magnitude is determined by the charging energy together with the random offset charges. However, we also allow for a n_s -independent energy term E_p that takes into account a possible contribution from polarons to the transport if the polaron relaxation time Δt exceeds $\Delta t > \hbar/E_p \approx 10^{-13}$ s. We note that this term is independent of the charge carrier density. The extrapolated tunnel resistance should be larger than the quantum resistance to be consistent with the model just described. We therefore associate with the quantum resistance a 'critical' mobility $\mu_c = e/(n_{\text{sites}}\hbar)$ which marks the boundary between thermally activated and temperature-independent transport if the intrinsic mobility is smaller ($\mu_0 \ll \mu_c$) or larger ($\mu_0 \gg \mu_c$) than the critical mobility.

The temperature dependent data of α -4T and α -6T single-crystal TFTs show a similar behaviour of the extrapolated conductivity and activation energy as a function of charge carrier density as the pentacene data already shown (Fig. 4). The transport processes of these highly ordered materials appear to be very similar, and show a strong dependence on the size of the molecule, which confirms the arguments of the orthodox Coulomb-blockade model.

In Table 1 we give the relevant parameters for the different materials. The tunnel barriers for these materials are indeed sufficiently large to be in the orthodox regime ($R_T \gg R_Q$). Note that the values of the intrinsic mobility μ_0 are smaller than the 'critical' mobility μ_c .

The threshold voltages in Table 1 are representative for the onset of p-type conduction for the different materials. The differences between the observed threshold voltage values are probably related to the work-function differences between the materials. However, the work-function difference between α -4T and α -6T is about 0.5 V (S. C. Veenstra *et al.*, manuscript in preparation), which is substantially smaller than the observed threshold voltage difference of 3.0 V. The exact origin of this effect is not understood. The listed values for the charging energies are rough estimates obtained from the maximum activation energy with an assumed number of random offset charges corresponding to an induced charge of $0.5e$ on the molecule. Alternatively, the charging energy can be calculated based on the self-capacitance of a α -4T molecule (assuming a sphere of radius r ($r = 16.4/2$ Å) the capacitance is about $C \approx 9 \times 10^{-19}$ F) which results in $E_c \approx 0.9$ eV.

All TFTs were fabricated using identical SiO_2 gate dielectric, and therefore we assume that the number of random offset charges in the SiO_2 is constant for all three materials. The observed activation energies should then directly scale with the charging energies and thus the capacitances of the molecules used, which to a first approximation is determined by their geometrical dimensions. This is demonstrated by the dependence of the observed activation

energies on charge carrier density. The scaling of the activation energies at identical charge carrier densities for α -4T and α -6T yields a constant value of 2, which can roughly be compared to a factor of 1.5 obtained with the self-capacitance (using the length) of both molecules. To improve the correspondence between both values, a more detailed approximation of the molecular capacitance, in which the actual charge distribution in relation to the nature and number of the charge carriers on the molecule, has to be taken into account. The activation energies do not tend to saturate at high carrier densities (see Fig. 4), which in our view confirms that polaron effects do not dominate transport.

We have performed numerical calculations in the orthodox regime, based on a two-dimensional array of 16×16 sites^{20,21}, similar to Fig. 1b, in which the dynamic properties are described by single-electron processes. A thermally activated transport behaviour is observed with activation energies dependent on charge carrier density. The magnitude is determined by the charging energy and number of random offset charges (data points in Fig. 4b). The simulations confirm that the extrapolation to infinite temperatures indeed yields the nearest-neighbour tunnel resistance. More detailed calculations of the molecular capacitance and the array need to be done in order to tighten the comparison with experiments.

The conductivity of organic TFTs is thus the result of an interplay between the intrinsic nearest-neighbour tunnel resistance and the capacitance of the individual molecules. Molecules with a higher capacitance will result in a lower activation energy and an increased conductivity. An alternative, more appealing, route to obtain organic materials with a high intrinsic mobility for device applications would be to decrease the tunnel resistance to the order of, or lower than, the quantum resistance (which is almost the case for α -4T). In such a regime, the orthodox Coulomb-blockade model breaks down, and charging effects will cease to play a role in transport, which would lead to high, temperature-independent mobilities ($\mu_0 > 0.5 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$). □

Table 1 Charge transport parameters of α -4T, α -6T and pentacene TFTs

Parameter	Material		
	α -4T	α -6T	Pentacene
V_t (V)	+2.0	+5.0	+9.6
n_{sites} (10^{18} m^{-2})	4.18	4.23	4.15
R_T (k Ω)	64	106	250
E_c (eV)	0.8	0.6	0.4
μ_0 ($10^{-1} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$)	2.3	1.4	0.6
μ_c ($10^{-1} \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$)	5.8	5.7	5.8

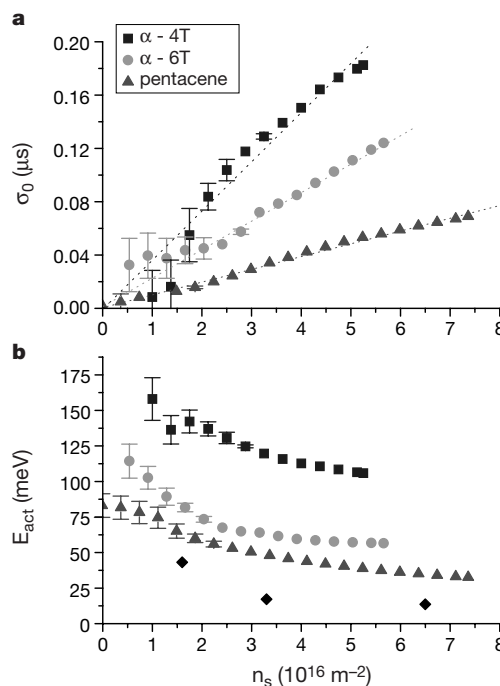


Figure 4 Comparison of the charge transport data of α -4T, α -6T and pentacene. **a**, The extrapolated conductivity σ_0 , and **b**, the activation energy, as a function of charge carrier density. The activation obtained from numerical simulations are plotted (diamonds) for carrier densities of 1, 2 and 4 charges in the two-dimensional array with $E_c = 0.4$ eV and random charges of 0.1 e.

Received 2 July 1999; accepted 3 February 2000.

- Grabert, H. & Devoret, M. H. in *Single Charge Tunneling* (eds Grabert, H. & Devoret, M. H.) Ch. 1 (NATO ASI Ser., Plenum, New York, 1992).
- Mooij, J. E. *et al.* Unbinding of charge-anticharge pairs in 2-dimensional arrays of small tunnel-junctions. *Phys. Rev. Lett.* **65**, 645–649 (1990).
- Horowitz, G., Hajlaoui, R. & Delannoy, P. Temperature-dependence of the field-effect mobility of sexithiophene—determination of the density of traps. *J. Phys. III* **5**, 355–371 (1995).
- Nelson, S. F., Lin, Y.-Y., Gundlach, D. J. & Jackson, T. N. Temperature-independent transport in high-mobility pentacene transistors. *Appl. Phys. Lett.* **72**, 1854–1856 (1998).
- Schön, J. H. & Batlogg, B. Modeling of the temperature dependence of the field-effect mobility in thin film devices of conjugated oligomers. *Appl. Phys. Lett.* **74**, 260–262 (1999).
- Waragai, K., Akimichi, H., Hotta, S. & Kano, H. Charge transport in thin-films of semiconducting oligothiophenes. *Phys. Rev. B* **52**, 1786–1792 (1995).
- Torsi, L., Dodabalapur, A., Rothberg, L. J., Fung, A. W. P. & Katz, H. E. Intrinsic transport properties and performance limits of organic field-effect transistors. *Science* **272**, 1462–1464 (1996).
- Torsi, L., Dodabalapur, A., Rothberg, L. J., Fung, A. W. P. & Katz, H. E. Charge transport in oligothiophene field-effect transistors. *Phys. Rev. B* **57**, 2271–2275 (1998).
- Vrijmoeth, J., Stok, R. W., Veldman, R., Schoonveld, W. A. & Klapwijk, T. M. Single crystallites in “planar polycrystalline” oligothiophene films: Determination of orientation and thickness by polarization microscopy. *J. Appl. Phys.* **83**, 3816–3824 (1998).
- Bouchoms, I. P. M., Schoonveld, W. A., Vrijmoeth, J. & Klapwijk, T. M. Morphology identification of the thin film phases of vacuum evaporated pentacene on SiO₂ substrates. *Synth. Met.* **104**, 175–178 (1999).
- Schoonveld, W. A., Vrijmoeth, J. & Klapwijk, T. M. Intrinsic charge transport properties of an organic single crystal determined using a multiterminal thin-film transistor. *Appl. Phys. Lett.* **73**, 3884–3886 (1998).
- Brown, A. R., Jarrett, C. P., De Leeuw, D. M. & Matters, M. Field-effect transistors made from solution-processed organic semiconductors. *Synth. Met.* **88**, 37–55 (1997).
- Vissenberg, M. C. J. M. & Matters, M. Theory of the field-effect mobility in amorphous organic transistors. *Phys. Rev. B* **57**, 12964–12967 (1998).
- Schoeler, H. in *Mesoscopic Electron Transport* (eds Sohn, L. L., Kouwenhoven, L. P. & Schön, G.) Ch. 1 (NATO ASI Ser., Kluwer Academic, Dordrecht, 1997).
- Porzio, W., Destri, S., Mascherpa, M. & Brückner, S. Structural aspects of oligothiophene series from X-ray powder diffraction data. *Acta. Polymer.* **44**, 266–272 (1993).
- Horowitz, G. *et al.* Growth and characterization of sexithiophene single-crystals. *Chem. Mater.* **7**, 1337–1341 (1995).
- Campbell, R. B., Robertson, J. M. & Trotter, J. The crystal structure of hexacene, and a revision of the crystallographic data for tetracene. *Acta Crystallogr.* **15**, 289–299 (1962).
- Holstein, T. Studies of polaron motion Part I. The molecular-crystal model. *Ann. Phys.* **8**, 325–387 (1959).
- Silins, E. A. & Čápek, V. *Organic Molecular Crystals* (AIP Press, New York, 1994).
- Geigenmüller, U. & Schön, G. Single-electron effects in arrays of normal tunnel-junctions. *Europhys. Lett.* **10**, 765–770 (1989).
- Bobbert, P. A., Geigenmüller, U., Fazio, R. & Schön, G. in *Macroscopic Quantum Phenomena* (eds Clark, T. *et al.*) 119–122 (World Scientific, New York, 1990).

Acknowledgements

We thank J. Vrijmoeth for his stimulating and crucial role in the early stages of this work, M. Mulder for support in sample fabrication and D. B. A. Rep for contributions to finalizing the manuscript. We also acknowledge encouragement and support from G. Hadzioannou. This work was supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) through the Strichting voor Fundamenteel Onderzoek der Materie (FOM).

Correspondence and requests for materials should be addressed to W.A.S. (e-mail: alex.schoonveld@philips.com).

Inorganic yellow-red pigments without toxic metals

M. Jansen* & H. P. Letschert†

* Max-Planck-Institut für Festkörperforschung, Heisenbergstrasse 1, D-70569 Stuttgart, Germany
† dmc², Degussa Metals Catalysts Cerdec, Postfach 11 04 03, D-60039 Frankfurt am Main, Germany

Inorganic pigments have been utilized by mankind since ancient times¹, and are still widely used to colour materials exposed to elevated temperatures during processing or application². Indeed, in the case of glasses, glazes and ceramics, there is no alternative to inorganic pigments for colouring. However, most inorganic pigments contain heavy metals or transition metals that can adversely effect the environment and human health if critical

levels are exceeded. Cadmium-based pigments in particular are a cause of concern³: although the pigments are not toxic due to their very low solubility in water and dilute mineral acids, cadmium itself is toxic and can enter the environment in a bioavailable form through waste-disposal sites and incineration plants⁴. This has led to regulations, based on the precautionary principle, that strongly restrict the use of cadmium pigments⁵. And even though recent assessments^{20,21} have concluded that the risk to humans or the environment might be not as significant as originally feared, a strong demand for inherently safer substitutes remains. Here we demonstrate that solid solutions of the perovskites CaTaO₂N and LaTaON₂ constitute promising candidates for such substitutes: their brilliance, tinting strength, opacity, dispersability, light-fastness and heat stability rival that of the cadmium pigments, while their colour can be tuned through the desired range, from yellow through orange to deep red, by simple composition adjustments. Because all the constituent elements are harmless, this perovskite-based inorganic pigment system seems a promising replacement that could eliminate one of the sources for cadmium emissions to the environment and some of the remaining concerns about pigment safety.

In general, colours of solids appear brilliant and pure when the corresponding mechanism for a selective absorption of light is related to an electronic interband transition, leading to a steep absorption edge in the visible spectrum. The width of the bandgap is determined by the extent of overlap of the valence orbitals, and by the difference between the electronegativities of the cations and anions involved. A concept^{6,7} has been derived from these factors (that is, valence-orbital overlap and electronegativity difference) that allows *a priori* design of bandgaps in semiconductors. In trying to rationally develop new inorganic pigments, we have adopted these ideas, mainly focusing on the option of widening the bandgap by increasing the difference of electronegativities between the respective cationic and anionic elements, and vice versa. Because we study optical effects, we have preferred to use the so-called optical electronegativities⁸. Whereas the electronegativities of the metals only vary in small steps, the substitution of one non-metal by another usually produces large shifts in electronegativity difference. Thus, in order to achieve a fine-tuning of bandgaps, we have focused on materials containing two different non-metals. For such compounds, the weighted average of the electronegativities of the respective anions has proved to be a good first approximation^{6,7}.

Table 1 Colour data of (Ca, La)TaO₂N₂ and of some commercial pigments

System	L*	a*	b*
Ca _(1-x) La _x TaO ₂ N ₂			
x = 0.05	77.70	-7.07	83.86
x = 0.15	62.34	9.82	80.50
x = 0.3	59.20	22.91	78.79
x = 0.45	52.16	28.64	73.58
x = 0.6	47.60	37.26	67.00
x = 0.75	35.21	42.35	55.71
x = 0.9	31.62	49.29	39.29
x = 1.0	26.35	37.01	30.08
Cadmium yellow	75.66	-7.90	99.12
Cadmium orange	64.68	47.70	96.26
Cadmium red	41.10	65.21	56.84
Cadmium dark Red	28.06	56.11	35.29
Fe ₂ O ₃	39.71	26.33	16.86
(Al _{0.95} Mn _{0.05}) ₂ O ₃	22.83	19.14	15.38

Shown are the colour coordinates of Ca_(1-x)La_xTaO₂N₂, x = 0.05–1.0, of four cadmium sulphoselenides and of two commercial red pigments without toxic metals. CIE Lab colour coordinates (L* = brightness axis, maximum values: 100 = white, 0 = black; a* = green-red axis, negative direction = green, positive direction = red; b* = blue-yellow axis, negative direction = blue, positive direction = yellow) determined according to DIN 5033, Part 3, 1976, 2° Observer, Illuminant C, 26% pigment in PVC. Commercial cadmium-yellow (Product No. 16308), -orange (Product No. 16309), -red (Product No. 18373), -dark red (Product No. 16176) from producer James M. Brown Limited, Stoke-on-Trent ST4 4NX, UK. Fe₂O₃ (Product No. RD101) and (Al_{0.95}Mn_{0.05})₂O₃ (Product No. FLO 172/411) from dmc² AG, Frankfurt/Main, Germany.



Most metal oxides have bandgaps too wide to allow for absorption in the visible region of the electromagnetic spectrum. By introducing nitrogen as a non-metal with lower electronegativity, the gap energy can be reduced by as much as ~ 1 eV, which makes oxonitrides promising candidates for inorganic pigments. In fact, some oxonitrides have already been found to be coloured^{9–11}. Next to producing the desired colour, another crucial requirement to be met is high chemical and physical stability. Taking into account the facts that the elements which form the most stable polar nitrides are tantalum, the alkaline-earth and the rare-earth elements, and that the perovskite type of structure usually represents a good matrix for durable solids, we are led almost directly to $ATa(O,N)_3$ compounds (where A is an alkaline-earth or rare-earth element) as potentially suitable systems.

In the course of a systematic search for perovskites that exhibit promising dielectric properties, some representatives of this family of compounds have been previously synthesized^{10–14}. Among these were $CaTaO_2N$ (ref. 10) and $LaTaON_2$ (refs 12,15), which mark the end members of the systems that we have selected to verify our concept of bandgap tuning as a rational tool for the design of

pigments. By applying an improved preparative approach, we have been able to show that the colours of these perovskites, if prepared in a pure state, are brilliant yellow and red, respectively. In fact, for $CaTaO_2N$ and $LaTaON_2$ complete mutual solubility exists¹⁶, which is particularly important, as it allows for a fine-tuning of the O/N ratio. According to the results of X-ray powder investigations¹⁷, all these solid solutions crystallize as tetragonally or orthorhombically distorted perovskites. It has been shown by neutron powder diffraction that the end members of the solid solutions contain oxygen and nitrogen in an ordered distribution¹⁷.

All samples, including the solid solutions, have brilliant colours, varying from bright yellow to deep red, depending on the O/N ratio. Ultraviolet–visible spectra accordingly feature steep absorption edges. For $Ca_{(1-x)}La_xTaO_{(2-x)}N_{(1+x)}$, the positions of the absorption edges vary linearly with x from 2.75 to 2.00 eV. Thus, any colour in the range from light yellow to deep red can be tailored by proper adjustment of the O/N ratio. In Table 1, the CIELab values of $Ca_{(1-x)}La_xTaO_{(2-x)}N_{(1+x)}$, and for comparison those of some commercial cadmium sulphoselenide pigments, are given, while Fig. 1 offers a graphic presentation within the a^*b^* colour space (see Table 1 legend and Fig. 1 legend for details of these colour description systems). As can be seen from these data, the $(Ca,L)Ta(O,N)_3$ pigments indeed fully match the yellow to red colour range of $Cd(S,Se)$, and the brilliance achieved by now is almost as good as exhibited by the highly optimized cadmium sulphoselenides. The most noticeable difference occurs for the end member $LaTaON_2$, which shows the same position on the energy scale, but a less steep absorption edge, compared to cadmium red (Fig. 2). This drawback does not appear to be intrinsic, as we have been able to steadily improve the colour performance of the lanthanum-rich pigments by optimizing the synthesis. This can most easily be achieved for the calcium-rich region, but becomes more challenging on increasing

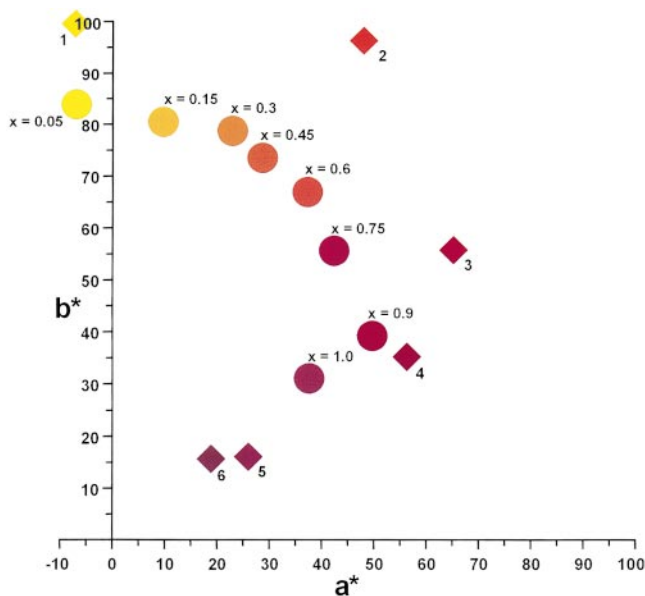


Figure 1 Pigment colour coordinates, with objective colours, in the $L^*a^*b^*$ space. The $Ca_{(1-x)}La_xTaO_{(2-x)}N_{(1+x)}$ pigments are compared to commercial cadmium yellow (1), cadmium orange (2), cadmium red (3), and cadmium dark red (4), and to some other reds without toxic metals, Fe_2O_3 (5) and $(Al_{0.95}Mn_{0.05})_2O_3$ (6). The $Ca_{(1-x)}La_xTaO_{(2-x)}N_{(1+x)}$ pigments are depicted as filled circles labelled by the corresponding x -values, and the cadmium sulphoselenides and the red pigments as filled diamonds labelled by the corresponding number, respectively. The CIELab colour coordinates are defined in Table 1 legend.

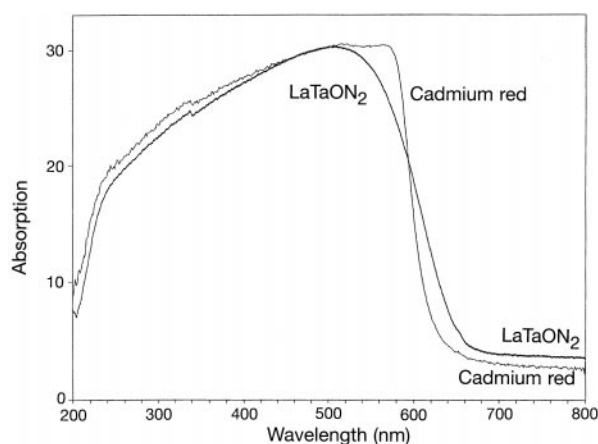


Figure 2 Comparison of the visible absorption spectra of the semiconductor pigments cadmium red and $LaTaON_2$.

Table 2 Comparison of colour values of cadmium yellow and $Ca_{(1-x)}La_xTaO_{(2-x)}N_{(1+x)}$ ($x = 0.05$)

Cadmium yellow								$Ca_{(1-x)}La_xTaO_{(2-x)}N_{(1+x)}$ ($x = 0.05$)							
L^*	a^*	b^*	C	h	O (%)	T (°C)	t (min)	L^*	a^*	b^*	C	h	O (%)	Ts (%)	
72.78	-5.35	73.46	73.65	94.17	69.60	600	4	73.73	-7.09	76.77	77.10	95.28	59.60	129.22	
72.88	-6.24	72.93	73.20	94.89	70.22	620	4	73.08	-7.34	77.31	77.66	95.42	59.57	146.28	
73.26	-8.03	68.24	68.71	96.71	71.88	640	4	72.06	-7.38	76.71	77.06	95.50	59.33	148.30	
74.36	-9.22	65.32	65.97	98.03	74.07	660	4	72.35	-7.29	76.75	77.10	95.43	61.67	169.84	
75.77	-9.70	61.71	62.47	98.93	76.90	660	6	71.55	-8.03	71.40	71.85	96.42	65.88	202.17	
76.38	-10.28	59.25	60.14	99.84	79.10	680	4	71.84	-7.91	73.65	74.07	96.13	64.27	210.91	

These colour values were obtained with the pigments in a glass paint flux. CIELab coordinates, as defined in Table 1; $C/h = L^*a^*b^*$ values in polar coordinates (C = Chroma: vector length in a^*/b^* plane, h = hue: vector angle from the a^* axis), O = opacity, Ts = tinting strength referenced to cadmium yellow (100%), T and t are the processing temperatures and times, respectively.

the lanthanum, and nitrogen, content. The reason is the substitution of divalent (Ca,O) for trivalent (La,N) constituents, which slows down the mobility in the solid state.

The thermal stability of these oxonitrides in an inert atmosphere (onset of mass loss for $\text{Ca}_{1.95}\text{La}_{0.05}\text{TaO}_{2.95}\text{N}_{1.05}$ is at 1,320 °C) is much better than that of the competing cadmium yellow (790 °C). The oxidation temperatures (onset) have been found to be virtually identical (420 °C) for both the yellow pigments under comparison. During the short period of processing to give lead-free enamels, the oxonitrides can be exposed to temperatures as high as 720 °C, while the upper limit for cadmium sulphoselenides lies at 650 °C. With the exception of boiling sulphuric acid, the new pigments resist the attack of solvents like HCl (25%), HNO_3 (65%), aqua regia, H_2O_2 (30%) and H_3PO_4 , at ambient and boiling temperatures.

The pigments^{15,16} presented here thus feature a unique combination of positive properties. They cover a broad section of the colour spectrum from yellow to deep red, are bright and clean in masstone (undiluted pigment), and have high tinting strength and opacity. These colorimetric properties, together with good dispersability, light-fastness and high thermal stability make them competitive with, or even superior to, the conventional cadmium sulphoselenide pigments. We confirmed this experimentally by comparing the pigmentary performances in glass paints of two yellow pigments, one each from the oxonitride and sulphoselenide systems (Table 2). While the opacity of the oxonitride is slightly less, the thermal stability in the glass flux, and in particular the tinting strength is clearly superior, compared to the sulphoselenide.

As demonstrated by laboratory tests, the field of possible applications covers colouring plastics, especially high-melting ones, and glass paints. For applications that involve processing at higher temperatures (for example, colouring glazes or ceramics), the pigment particles will have to be included into, or coated by, refractive materials like ZrSiO_4 , or ZrO_2 (refs 18, 19). □

Methods

Chemical syntheses

The polycrystalline samples studied here were made by solid-state synthesis. CaCO_3 (p.a., Riedel-de Haen), La_2O_3 and Ta_2O_5 aq. in the required stoichiometric ratios—with CaCl_2 , KCl and NaCl in equimolar quantities added as mineralizers—were mixed by wet milling (ethanol) and reacted at 850 °C for 20–60 h in a flow of dry ammonia. The 1.5-g batches contained starting materials and mineralizers in about 50 wt% each. The reactive starting materials La_2O_3 and Ta_2O_5 aq. were made by heating $\text{La}(\text{NO}_3)_3 \cdot 5\text{H}_2\text{O}$ (Aldrich) *in vacuo* at 500 °C, and by hydrolysing TaCl_5 , respectively. After leaching the halides with water, single-phase products were obtained.

Analytical characterization

The samples were characterized by X-ray and neutron powder diffraction, thermal analysis (Netsch STA 409, heating rate 5 K min⁻¹) and UV–vis spectroscopy (Varian Cary 2400; 100 mg sample in 4.5 g BaSO_4 , diffuse reflection mode).

Colour measurements

All pigmentary data were obtained using the Hunterlab Labscan 5100 (Illuminant C, 2° standard observer), and applying the prescriptions fixed as International Standard ISO 787, part 24 for the relative tinting strength²², and ISO 247 for the opacity²³.

The comparison of the performance as glass colours was executed for a pair of yellow pigments, cadmium yellow (dmc^2) and $\text{Ca}_{0.95}\text{La}_{0.05}\text{TaO}_{1.95}\text{N}_{1.05}$ using a glaze of composition (wt%) 41% SiO_2 , 20% ZnO , 13% Na_2O , 3% K_2O , 12% B_2O_3 , 8% TiO_2 and 3% Al_2O_3 . Mixtures consisting of 5 wt% pigment and 95 wt% glaze were deposited onto glass plates by screen-printing. The burning temperatures and durations applied are given in Table 2.

Received 27 July 1999; accepted 13 March 2000.

- Noll, W., Reimer, H. & Born, L. Bemalung antiker Keramik. *Angew. Chem.* **87**, 639–651 (1975).
- Buxbaum, G. *Industrial Inorganic Pigments* (VCH, Weinheim, 1993).
- Wirth, W. & Gloxhuber, C. *Toxikologie* (Thieme, Stuttgart, 1981).
- Knorn, D. Cadmiumemissionen bei der Verbrennung und Ablagerung von Hausmüll. *Metall* **37**, 633–635, (1983).
- European Economic Community Guideline No. 91/338/EWG (1991).
- Van Vechten J. A. & Phillips, J. C. New set of tetrahedral covalent radii. *Phys. Rev. B* **2**, 2160–2167 (1970).
- Phillips, J. C. *Bonds and Bands in Semiconductors* (Academic, New York, 1973).
- Jørgensen, C. K. *Oxidation Numbers and Oxidation States* (Springer, Berlin, 1969).

- Strähle, J. Die Kristallstruktur des Tantal(V)-nitrides. *Z. Anorg. Allg. Chem.* **402**, 47–57 (1973).
- Marchand, R., Pors, F. & Laurent, Y. Préparation et caractérisation de nouveaux oxynitrides à structure perovskite. *Rev. Int. Hautes Tempér. Réfract.* **23**, 11–15 (1986).
- Grins, J. & Svensson, G. Synthesis of oxynitride perovskites $[\text{AZr}_x\text{Ta}_{1-x}\text{O}_{2+x}\text{N}_{1-x}]$ A = Ca, Sr, Ba and 0 < x < 1. *Mater. Res. Bull.* **29**, 801–809 (1994).
- Marchand, R., Pors, F. & Laurent, Y. Nouvelles pérovskites oxynitrides de stoechiométrie ABO_2N (A = Lanthanide, B = Ti) et ABON_2 (A = Lanthanide B = Ta ou Nb). *Ann. Chim. Fr.* **16**, 553–560 (1991).
- Marchand, R., Laurent, Y., Guyader, J., L'Haridon, P. & Verdier, P. Nitrides and oxynitrides: Preparation, crystal chemistry and properties. *J. Eur. Ceram. Soc.* **8**, 197–213 (1991).
- Pors, F., Marchand, R., Laurent, Y., Bacher, P. & Rault, G. Étude structurale des pérovskites oxyazotes BaNbO_2N et BaTaO_2N . *Mater. Res. Bull.* **23**, 1447–1450 (1988).
- Jansen, M., Letschert, H. P. & Speer, D. Oxonitrides of the formula LnTaON_2 with enhanced brightness and a process for their use. European Patent No. 0627382 (1992).
- Jansen, M. & Letschert, H. P. Oxynitride-based coloured pigments and process for their preparation. European Patent No. 0697373 (1995).
- Günther, E., Hagemeyer, R. & Jansen, M. Strukturuntersuchungen an den Oxidnitriden SrTaO_2N , CaTaO_2N und LaTaON_2 mittels Neutronen- und Röntgenbeugung. *Z. Anorg. Allg. Chem.* (in the press).
- Beyer, H. H., Broll, A. & Kleinschmit, P. Einschlusßpigmente, Farbkörper nach einem neuen Prinzip. *Chemiker-Zeitung* **101**, 319–323 (1977).
- Beyer, H. H., Broll, A., Mann, H. & Meyer-Simon, E. P. Keramische Farbkörper. German Patent DE2312535 (1972).
- WS Atkins International Assessment of the risks of health and to the environment of Cadmium contained in certain products and of the effects of further restriction on their marketing and use (Surrey, UK, September 1998).
- http://europa.eu.int/comm/dg24/health/sc/out28_en.html
- International Standard ISO ISO 787 part 24 (Beuth, Berlin, 1986).
- International Standard ISO ISO 747 (Beuth, Berlin, 1988).

Correspondence and requests for materials should be addressed to M.J. (e-mail: martin@jansen.mpi-stuttgart.mpg.de).

A homochiral metal–organic porous material for enantioselective separation and catalysis

Jung Soo Seo, Dongmok Whang, Hyoyoung Lee, Sung Im Jun, Jinho Oh, Young Jin Jeon & Kimoon Kim

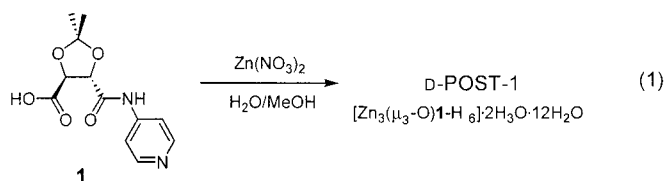
National Creative Research Initiative Center for Smart Supramolecules and Department of Chemistry, Division of Molecular and Life Sciences, Pohang University of Science and Technology, San 31 Hyojadong, Pohang 790-784, Republic of Korea

Inorganic zeolites are used for many practical applications that exploit the microporosity intrinsic to their crystal structures. Organic analogues, which are assembled from modular organic building blocks linked through non-covalent interactions, are of interest for similar applications. These range from catalysis, separation and sensor technology to optoelectronics^{1–3}, with enantioselective separation and catalysis being especially important for the chemical and pharmaceutical industries. The modular construction of these analogues allows flexible and rational design, as both the architecture and chemical functionality of the micropores can, in principle, be precisely controlled. Porous organic solids with large voids and high framework stability have been produced^{14,15}, and investigations into the range of accessible pore functionalities have been initiated^{7,11,12,16–22}. For example, catalytically active organic zeolite analogues are known^{13,22,23}, as are chiral metal–organic open-framework materials. However, the latter are only available as racemic mixtures^{24,25}, or lack the degree of framework stability or void space that is required for practical applications^{26,27}. Here we report the synthesis of a homochiral metal–organic porous material that allows the enantioselective inclusion of metal complexes in its pores and catalyses a transesterification reaction in an enantioselective manner. Our synthesis strategy, which uses enantiopure metal–organic clusters as secondary building blocks¹⁴, should be readily applicable to chemically modified cluster components and thus provide access

to a wide range of porous organic materials suitable for enantio-selective separation and catalysis.

Our strategy for achieving rigid, homochiral frameworks with large chiral voids exploits large, robust, enantiopure chiral metal–organic clusters as building blocks¹⁴. Potential candidates for such building blocks include oxo-bridged trinuclear metal carboxylates (often known as ‘basic’ carboxylates) $[M_3(\mu_3-O)(O_2CR)_6(H_2O)_3]^{n+}$ (M = divalent or trivalent transition metal ions, O_2CR = organic carboxylate anions) (Fig. 1), which are commonly found in transition metal coordination chemistry and are readily assembled from metal ions and carboxylates²⁸. The water molecules bound to the metal centres are easily replaced by a Lewis base such as pyridine. We therefore thought that a rigid chiral organic molecule with a carboxylic group at one end and a pyridyl group at the other might serve the dual function of forming oxo-bridged trinuclear metal carboxylate units and linking these trimers into two- or three-dimensional frameworks.

The enantiopure chiral organic building block **1**, which can be easily synthesized from D-tartaric acid, reacts with Zn^{2+} ions to produce a homochiral open-framework solid whose formula is given as $[Zn_3(\mu_3-O)(1-H)_6] \cdot 2H_2O \cdot 12H_2O$ (referred to as D-POST-1) (equation (1)). The enantiomorphous L-POST-1 is obtained from the enantiomer of **1** and the Zn^{2+} ion under the same conditions (see Supplementary Information).



In the crystal structure of D-POST-1, three zinc ions are held together with six carboxylate groups of the deprotonated chiral

ligands and a bridging oxo oxygen, to form a trinuclear unit (Fig. 2), in which a threefold axis parallel to the c axis passes through the centre of the trinuclear unit. The three zinc ions and the bridging oxo oxygen are located on the same plane. In POST-1, the trinuclear units, which behave as a secondary building unit¹⁴, are interconnected through coordinate covalent bonds between the zinc ions and pyridyl groups of **1**, thereby generating two-dimensional (2D) infinite layers consisting of large edge-sharing chair-shaped hexagons with a trinuclear unit at each corner (Figs 3 and 4). The 2D layers stack along the c axis with a mean interlayer separation of 15.47 Å. The structure is apparently stabilized by efficient van der Waals interactions between the 2D layers arising from their self-complementary structure. Most notably, large chiral 1D channels exist along the c axis, the cross-section of which is best described as an equilateral triangle with a side length of ~ 13.4 Å (Fig. 4). The void volume of the channels is estimated to be $\sim 47\%$ of the total volume and is filled with water molecules (47 water molecules per unit cell). POST-1 is stable in normal organic solvents, and slightly soluble only in water and dimethyl sulfoxide (DMSO). POST-1 loses crystallinity upon removal of the solvate molecules by evacuation, but restores it upon exposure of the evacuated sample to ethanol or water vapour as judged by powder X-ray diffraction patterns (see Supplementary Information). Further characterization of porosity, rigidity and stability of the framework remains to be done.

There are six pyridyl groups per trinuclear unit ($[Zn_3(\mu_3-O)(1-H)_6]^{2-}$) in the structure. Three of them are coordinated to the zinc ions of three neighbouring trinuclear units, the other three extrude into the channel without any interactions with the framework (Fig. 3). Considering that the zinc ion, the deprotonated ligand **1** and bridging oxo oxygen have charges of +2, –1 and –2, respectively, we need two additional positive charges per $[Zn_3(\mu_3-O)(1-H)_6]^{2-}$ unit for overall neutrality. Given that there are no other counter ions

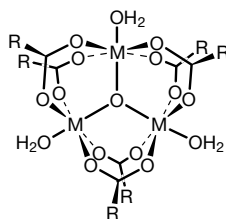


Figure 1 Oxo-bridged trinuclear metal carboxylates $[M_3(\mu_3-O)(O_2CR)_6(H_2O)_3]^{n+}$ (M = divalent or trivalent transition metal ions, O_2CR = organic carboxylate anions).

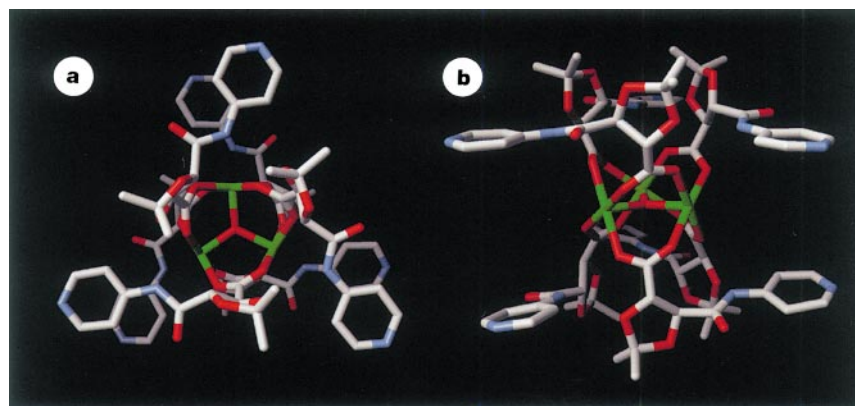


Figure 2 Structure of the trinuclear secondary building unit. The secondary building unit $[Zn_3(\mu_3-O)(1-H)_6]^{2-}$ is shown in top view (a) and side view (b). Each zinc ion is coordinated by the bridging oxo moiety, four carboxyl oxygen atoms of **1** and a pyridyl nitrogen atom of

a neighbouring trinuclear unit in a distorted octahedral geometry. Zn atoms are shown in green, O in red, N in blue and C in white.

found in the crystal structure, which has been confirmed by elemental analysis, it is reasonable to assume that the two positive charges required are provided by protonation of two of the three pyridyl groups exposed in the channel. Cation exchange experiments subsequently showed that the protons bound to the pyridyl groups can be exchanged with alkali metal ions, such as Na^+ , K^+ and Rb^+ . For example, soaking POST-1 in an ethanol solution of sodium *p*-toluenesulphonate at room temperature for 12 h results in almost complete exchange of the cation without structural change of the host framework.

Owing to the chiral environment of the channel, enantioselective inclusion of cationic metal complexes can also be achieved. Upon suspending L-POST-1 in a methanol solution of

racemic $[\text{Ru}(2,2'\text{-bipy})_3]\text{Cl}_2$ (bipy = bipyridine), the colour of L-POST-1 changes from white to reddish yellow. Nuclear magnetic resonance (NMR), ultraviolet–visible and circular dichroism measurements showed that 80% of the exchangeable protons are exchanged with $[\text{Ru}(2,2'\text{-bipy})_3]^{2+}$ (0.8 Ru complex per $[\text{Zn}_3(\mu_3\text{-O})(1\text{-H})_6]^{2-}$ unit) with 66% enantiomeric excess in favour of the Δ form (see Supplementary Information). Such enantioselective inclusion of chiral metal complexes in modular porous materials is unprecedented.

The pyridyl groups exposed in the channels endow POST-1 with unique opportunities for chemical modification of the channel environment (Fig. 3). For example, the *N*-alkylation of all the pyridyl groups can change the framework charge of POST-1 from

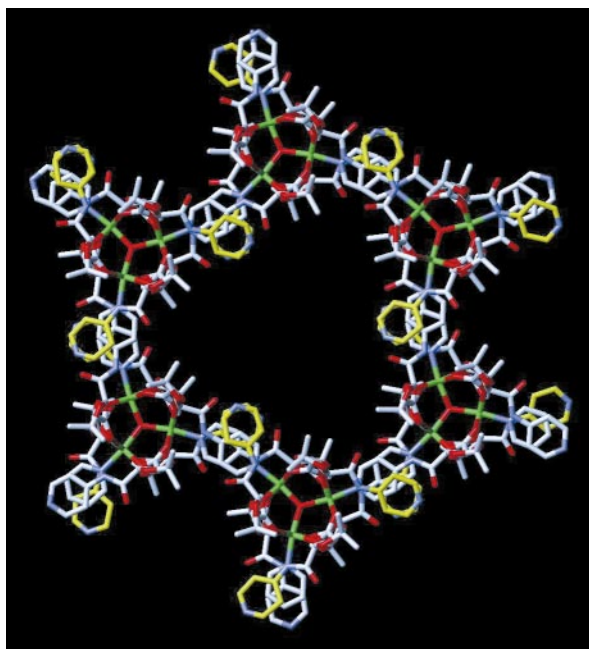


Figure 3 The hexagonal framework with large pores that is formed with the trinuclear secondary building units. The colours are as in Fig. 2 except for C atoms in the pyridyl groups that are exposed in the pore which are shown in yellow.

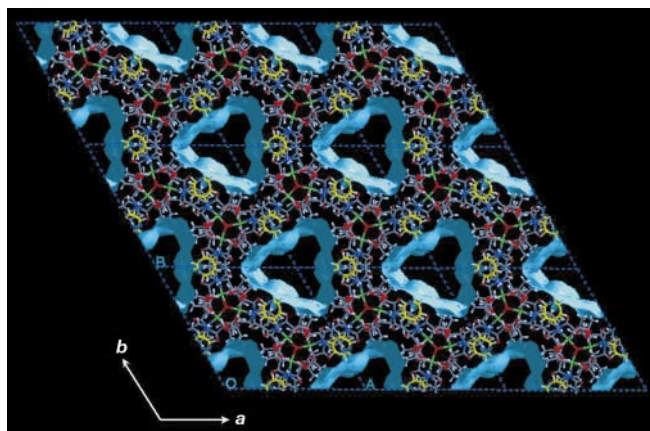


Figure 4 View down the *c* axis showing large chiral channels of POST-1. Zn ions (green), oxo moieties (red) and the pyridyl (py) groups exposed in the channels (yellow) are highlighted. The accessible surface in the channels is shown in blue. Properties of POST-1 are as follows. Crystal data for D-POST-1: trigonal, space group $P321$ with $a = 20.9415(6) \text{ \AA}$, $c = 15.4720(6) \text{ \AA}$, $V = 5,876.1(3) \text{ \AA}^3$ and $Z = 2$ (for empirical formula $\text{C}_{72}\text{H}_{98}\text{N}_{12}\text{O}_{40}\text{Zn}_3$). $T = 185(2) \text{ K}$, R_1 (unweighted, based on F^2) = 0.102. Solid

state ^{13}C cross polarization magic-angle spinning NMR: δ 174.5 (s, $-\text{COO}-$); 168.1 (s, $-\text{CONHpy}$); 149.2 (s, $-\text{pyC}_0$); 113.0 (s, $-\text{pyC}_0$); 79.2 (s, $(\text{CH}_3)_2\text{C}(\text{O})_2(\text{CHCOO}^-)-(\text{CHCONHpy})$); 26.4 (s, $(\text{CH}_3)_2\text{C}=\text{C}$). Elemental microanalysis: calculated (%) for $[\text{Zn}_3(\text{O})(1\text{-H})_6](\text{H}_3\text{O})_2(\text{H}_2\text{O})_7$: C, 43.95; H, 5.02; N, 8.54; Zn, 9.97. Found (%): C, 43.66; H, 4.93; N, 8.64; Zn, 10.27.

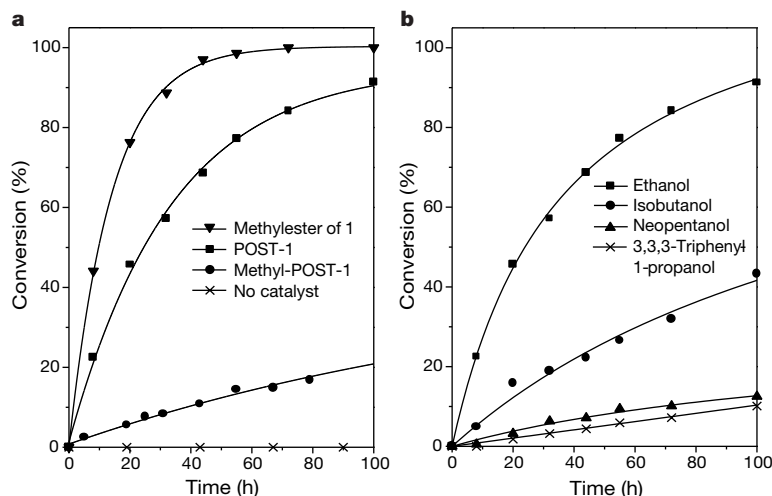


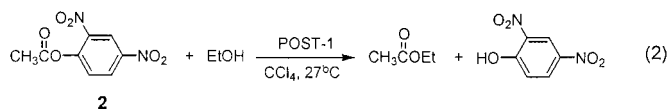
Figure 5 Catalytic activity of POST-1 in transesterification reactions. **a**, Transesterification of **2** with ethanol in the presence of POST-1. The reaction was carried out in carbon tetrachloride at 27 °C in the absence of catalyst, and the presence of 0.1 equivalents (based on the number of pyridyl units exposed to channels) of POST-1, methylated POST-

1 and the methyl ester of **1** (**2**:EtOH = 1:4). **b**, Transesterification of **2** with alcohols in the presence of POST-1. The reaction was carried out with ethanol, isobutanol, neopentanol or 3,3,3-triphenyl-1-propanol (**2**:alcohol = 1:4) in carbon tetrachloride at 27 °C in the presence of 0.1 equivalents of POST-1.

negative to positive. Treatment of a suspension of crystalline POST-1 in *N,N'*-dimethylformamide (DMF) with excess iodomethane at room temperature for 2 h results in the conversion of all the pyridyl groups exposed to the channels to *N*-methyl pyridinium ions as indicated by ¹H NMR analysis. However, powder X-ray analysis reveals the framework structure to be unchanged. Raman spectroscopy confirmed the presence of I₃⁻ counterions in the methylated POST-1. The anion can be reversibly exchanged with other anions. For example, the addition of a slight excess of NaPF₆ to a suspension of the methylated POST-1 in DMF results in full exchange of I₃⁻ with PF₆⁻ in the modified POST-1, as determined by infrared spectroscopy. The exchanged PF₆⁻ can be in turn exchanged with *p*-toluenesulphonate (CH₃C₆H₄SO₃⁻).

The pore size of POST-1 is also modulated by *N*-alkylation of the pyridyl groups. Thermal analysis of the modified porous solids, the channels of which had been filled with DMF molecules, showed that the pore volume of POST-1 shrinks by 14% and 60% upon *N*-alkylation with iodomethane (CH₃I) and 1-iodohexane (CH₃(CH₂)₅I), respectively (see Supplementary Information). The amount of anions that can be included in the channel also decreases as the pore volume shrinks.

The presence of the pyridyl groups exposed in the channels also provides POST-1 with unique opportunities in catalysis. We examined the catalytic activity of POST-1 for transesterification of ester **2** (equation (2)). Treatment of **2** and ethanol with suspension of POST-1 (0.1 equivalents) in carbon tetrachloride for 55 h at 27 °C produced ethyl acetate in 77% yield. No or little transesterification occurs without POST-1 or with the *N*-methylated POST-1, respectively (Fig. 5a). Transesterification of **2** with bulkier alcohols such as isobutanol, neopentanol and 3,3,3-triphenyl-1-propanol occurs with a much slower or even negligible rate under otherwise identical reaction conditions (Fig. 5b). Such size selectivity suggests that the catalysis mainly occurs in the channels.



POST-1 revealed enantioselective catalytic activity for the transesterification of **2**. A preliminary study showed that the reaction of **2** with a large excess of racemic 1-phenyl-2-propanol in the presence of D-POST-1 or the enantiomeric L-POST-1 produces the corre-

sponding esters with ~8% enantiomeric excess in favour of *S* or *R* enantiomer, respectively. Although modest, the enantioselectivity of this process is noteworthy because asymmetric induction has never been observed in reactions mediated by modular porous materials. Further studies on enantioselective catalytic activities of POST-1, including a search for reactions and substrates that may lead to higher enantiomeric excess values, are currently underway. We anticipate that this design strategy, applied to chemically modified starting materials, will allow the introduction of other recognition and/or catalytic sites into the channels of analogous porous materials, which may find useful applications. □

Received 21 December 1999; accepted 13 March 2000.

- Yaghi, O. M., Li, H., Davis, C., Richardson, D. & Groy, T. L. Synthetic strategies, structure patterns, and emerging properties in the chemistry of modular porous solids. *Acc. Chem. Res.* **31**, 474–484 (1998).
- Aoyama, Y. Functional organic zeolite analogues. *Top. Curr. Chem.* **198**, 131–161 (1998).
- Kitagawa, S. & Kondo, M. Functional micropore chemistry of crystalline metal complex-assembled compounds. *Bull. Chem. Soc. Jpn* **71**, 1739–1753 (1998).
- Janiak, C. Functional organic analogues of zeolites based on metal-organic coordination frameworks. *Angew. Chem. Int. Edn Engl.* **36**, 1431–1434 (1997).
- Abrahams, B. E., Hoskins, B. E., Michail, D. M. & Robson, R. Assembly of porphyrin building blocks into network structures with large channels. *Nature* **369**, 727–729 (1994).
- Yaghi, O. M., Li, G. & Li, H. Selective binding and removal of guests in a microporous metal-organic framework. *Nature* **378**, 703–706 (1995).
- Aoyama, Y. et al. Crystal engineering of stacked aromatic columns. Three-dimensional control of the alignment of orthogonal aromatic triads and guest quinones via self-assembly of hydrogen-bonded networks. *J. Am. Chem. Soc.* **118**, 5562–5571 (1996).
- Gardner, G. B., Venkataraman, D., Moore, J. S. & Lee, S. Spontaneous assembly of a hinged coordination network. *Nature* **374**, 792–795 (1995).
- Subramanian, S. & Zaworotko, M. J. Porous solids by design: [Zn(4,4'-bipy)₂(SiF₆)₂]_n·xDMF, a single framework octahedral coordination polymer with large square channels. *Angew. Chem. Int. Edn Engl.* **34**, 2127–2129 (1995).
- Russell, V. A., Evans, C. C., Li, W. & Ward, M. D. Nanoporous molecular sandwiches: pillared two-dimensional hydrogen-bonded networks with adjustable porosity. *Science* **276**, 575–579 (1997).
- Brunet, P., Simart, M. & Wuest, J. D. Molecular tectonics. Porous hydrogen-bonded networks with unprecedented structural integrity. *J. Am. Chem. Soc.* **119**, 2737–2738 (1997).
- Fujita, M., Kwon, Y. J., Washizu, S. & Ogura, K. Preparation, clathration ability, and catalysis of a two-dimensional square network material composed of cadmium(II) and 4,4'-bipyridine. *J. Am. Chem. Soc.* **116**, 1151–1152 (1994).
- Heo, J., Kim, S. Y., Whang, D. & Kim, K. Shape-induced, hexagonal, open frameworks: rubidium ion complexed cucurbituril. *Angew. Chem. Int. Edn* **38**, 641–643 (1999).
- Li, H., Eddaoudi, M., O'Keeffe, M. & Yaghi, O. M. Design and synthesis of an exceptionally stable and highly porous metal-organic framework. *Nature* **402**, 276–279 (1999).
- Chui, S. S.-Y., Lo, S. M.-F., Charmant, J. P. H., Orpen, A. G. & Williams, I. D. A chemically functionalizable nanoporous material [Cu₃(TMA)₂(H₂O)₃]_n. *Science* **283**, 1148–1150 (1999).
- Kiang, Y.-H., Gardner, G. B., Lee, S., Xu, Z. & Lobkovsky, E. B. Variable pore size, variable chemical functionality, and an example of reactivity within porous phenylacetylene silver salts. *J. Am. Chem. Soc.* **121**, 8204–8215 (1999).
- Gardner, G. B., Kiang, Y.-H., Lee, S., Asgaonkar, A. & Venkataraman, D. Exchange properties of the three-dimensional coordination compound 1,3,5-Tris(4-ethynylbenzonitrile)benzene-AgO₃SCF₃. *J. Am. Chem. Soc.* **118**, 6946–6953 (1996).

18. Lin, K.-J. SMTP-1: the first functionalized metalloporphyrin molecular sieves with large channels. *Angew. Chem. Int. Edn* **38**, 2730–2732 (1999).
19. Li, H., Davis, C. E., Groy, T. L., Kelley, D. G. & Yaghi, O. M. Coordinatively unsaturated metal centers in the extended porous framework of $Zn_3(\text{BDC})_2 \cdot 6\text{CH}_3\text{OH}$ (BDC = 1,4-Benzenedicarboxylate). *J. Am. Chem. Soc.* **120**, 2186–2187 (1998).
20. Reineke, T. M., Eddaoudi, M., Fehr, M., Kelley, D. & Yaghi, O. M. From condensed lanthanide coordination solids to microporous frameworks having accessible metal sites. *J. Am. Chem. Soc.* **121**, 1651–1657 (1999).
21. Kondo, K., Yoshitomi, T., Seki, K., Matsuzaka, H. & Kitagawa, S. Three-dimensional framework with channeling cavities for small molecules: $[\text{M}_2(4,4'\text{-bpy})_2(\text{NO}_3)_4] \cdot x\text{H}_2\text{O}$ ($\text{M} = \text{Co}, \text{Ni}, \text{Zn}$). *Angew. Chem. Int. Edn Engl.* **36**, 1725–1727 (1997).
22. Endo, K. *et al.* Catalysis by organic solids. Stereoselective Diels–Alder reactions promoted by microporous molecular crystals having an extensive hydrogen-bonded network. *J. Am. Chem. Soc.* **119**, 4117–4122 (1997).
23. Sawaki, T. & Aoyama, Y. Immobilization of a soluble metal complex in an organic network. Remarkable catalytic performance of a porous dialkoxymethylzirconium polyphenoxide as a functional organic zeolite analogue. *J. Am. Chem. Soc.* **121**, 4793–4798 (1999).
24. Biradha, K., Seward, C. & Zaworotko, M. J. Helical coordination polymers with large chiral cavities. *Angew. Chem. Int. Edn Engl.* **38**, 492–495 (1999).
25. Decurtins, S., Schmalke, H. W., Schneuwly, P., Ensling, J. & Güthlich, P. A concept for the synthesis of 3-dimensional homo- and bimetallic oxalate-bridged networks $[\text{M}_2(\text{ox})_3]_n$. Structural, Mössbauer, and magnetic studies in the field of molecular-based magnets. *J. Am. Chem. Soc.* **116**, 9521–9528 (1994).
26. Normar, R. E., Rose, N. J. & Stenkamp, R. E. Crystal structure of a copper complex of 2-carboxypentonic acid; a decomposition product of dehydroascorbic acid. *J. Chem. Soc. Dalton Trans.* 2905–2910 (1987).
27. Ranford, J. D., Vittal, J. J. & Wu, D. Topochemical conversion of a hydrogen-bonded three-dimensional network into a covalently bonded framework. *Angew. Chem. Int. Edn Engl.* **37**, 1114–1116 (1998).
28. Hathaway, B. J. in *Comprehensive Coordination Chemistry* Vol. 2 (ed. Wilkinson, G.) 439–441 (Pergamon, Oxford, 1987).

Supplementary information is available on Nature's World-Wide Web site or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank the Korean Ministry of Science and Technology (Creative Research Initiatives Program) for supporting this work and the Korean Ministry of Education (Brain Korea 21 program) for graduate student fellowships (J.O. and Y.J.J.). We also thank M. G. Finn and L. K. Woo for critical reading of the manuscript.

Correspondence and requests for materials should be addressed to K.K. (e-mail: kkim@postech.ac.kr).

Recycled oceanic crust observed in 'ghost plagioclase' within the source of Mauna Loa lavas

Alexander V. Sobolev^{*†}, Albrecht W. Hofmann^{*} & Igor K. Nikogosian^{*‡†}

^{*} Max-Planck-Institut für Chemie, Postfach 3060, 55020 Mainz, Germany

[†] Vernadsky Institute of Geochemistry, Russian Academy of Sciences, Kosygin Street 19, 117975 Moscow, Russia

[‡] Department of Petrology, Vrije Universiteit, De Boelelaan 1085, 1081 HV, Amsterdam, The Netherlands

The hypothesis that mantle plumes contain recycled oceanic crust¹ is now widely accepted. Some specific source components of the Hawaiian plume have been inferred to represent recycled oceanic basalts², pelagic sediments^{3,4} or oceanic gabbros⁵. Bulk lava compositions, however, retain the specific trace-element fingerprint of the original crustal component in only a highly attenuated form. Here we report the discovery of exotic, strontium-enriched melt inclusions in Mauna Loa olivines. Their complete trace-element patterns strongly resemble those of layered gabbros found in ophiolites, which are characterized by cumulus plagioclase with very high strontium abundances⁶. The major-element compositions of these melts indicate that their composition cannot be the result of the assimilation of present-day oceanic crust through which the melts have travelled. Instead,

the gabbro has been transformed into a (high-pressure) eclogite by subduction and recycling, and this eclogite has then been incorporated into the Hawaiian mantle plume. The trace-element signature of the original plagioclase is present only as a 'ghost' signature, which permits specific identification of the recycled rock type. The 'ghost plagioclase' trace-element signature demonstrates that the former gabbro can retain much of its original chemical identity through the convective cycle without completely mixing with other portions of the former oceanic crust.

We studied melt inclusions in olivine phenocrysts in three lavas covering 50 kyr of eruption history of Mauna Loa volcano, Hawaii. The lavas are picrites with more than 20% olivine phenocrysts. We examined 250 melt inclusions by electron microprobe, and selected 160 of these for ion microprobe analysis. All measured compositions were corrected for Fe–Mg exchange between included melt and host olivine, and for crystallization of olivine on the cavity walls (see Methods). They are believed to represent the primary compositions of entrapped melts.

The corrected major-element compositions of the melt inclusions are similar to primitive Mauna Loa lavas, and show the olivine control characteristics typical of Hawaiian tholeiites (Fig. 1). This observation and numerical simulation of plagioclase melting using mineral–melt geothermometry⁷ suggest that all reconstructed trapped melts are well within the olivine crystallization field at low pressures and are considerably undersaturated with respect to plagioclase at any pressure of its stability (Fig. 1).

The melt inclusions show significant variations in their incompatible-trace-element concentrations. We distinguish two groups of melts on the basis of Sr concentration and Sr/REE ratio (Figs 2 and 3; REE, rare-earth element). Most of the melt inclusions show moderate enrichment of Sr relative to Ce, and will be referred to as 'normal'. Six inclusions show extreme Sr excess, and will be called 'Sr-rich'.

The average trace-element composition of the inclusions is remarkably similar to the Mauna Loa lavas (Fig. 2a)⁵. However, the range of incompatible element (for example, light-REE, Ba, Th) concentrations and ratios is much greater than is found in Hawaiian

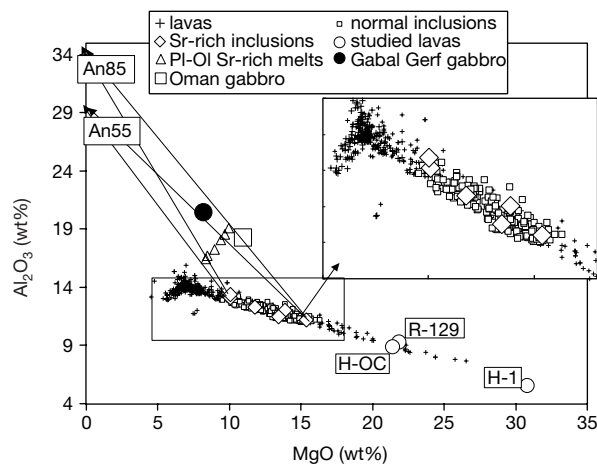


Figure 1 Al_2O_3 versus MgO in lavas and inclusions. The figure shows data for Mauna Loa lavas (see GEOROC database at <http://georoc.mpch-mainz.gwdg.de>), corrected normal and Sr-rich melt inclusions, bulk compositions of the three investigated lava samples (with sample numbers), and average compositions of ophiolite gabbros from Gabal Gerf⁶ and Oman (C. J. Garrido & P. B. Kelemen, manuscript in preparation). The open triangles show Sr-rich melts saturated in plagioclase and olivine, calculated by PETROLOG software⁷ (see Supplementary Information). Arrows show assimilation trends of plagioclase with different anorthite contents (An in mol%). The apparent olivine control trend is similar for all inclusion types and lavas. All lavas with $\text{MgO} > 7\%$ and all inclusions are plagioclase-undersaturated. Sr-rich inclusions follow the olivine control trend, and show no sign of Al enrichment resulting from interaction of melt with gabbro or plagioclase.

tholeiitic lavas. This is illustrated in Fig. 3, where $(\text{La}/\text{Sm})_n$ (here subscript n means normalization to the composition of primitive mantle³⁰) is used as a measure of the relative depletion or enrichment of incompatible elements. In normal inclusions, $(\text{La}/\text{Sm})_n$ varies from 0.3 to 1.9, whereas Mauna Loa lavas⁸ show a much narrower range of 1.0–1.6.

Less than 4% of the inclusions are Sr-rich. Compared with normal inclusions and Mauna Loa lavas, Sr-rich melts are also severely depleted in the most incompatible elements (Ba, Th, Nb, K, La, Ce, Nd and Zr). In contrast to the exotic incompatible-element signatures, heavy-REE and major-element compositions of these melts are typical for Mauna Loa lavas (Figs 1 and 3; Tables 1 and 2 in Supplementary Information). Two features of the heavy-REE distributions, their relative depletion and their low variability (within analytical error), indicate that garnet was in equilibrium with both

normal and Sr-rich melts.

In most of the olivine grains, the composition of the melt inclusions is fairly uniform. This has been demonstrated by closely spaced, sequential thin-sectioning, which yielded up to 21 separate inclusions in 6 sections of a single olivine grain. However, the exotic melt inclusions with either high Sr/Ce or extreme La/Sm ratios are invariably associated with normal inclusions in the same grain. The most extreme example is olivine grain HOC-76. Here, the concentrations of Ba, Nb and K differ by a factor of more than 10 in melt inclusions separated by a few hundred micrometres of host olivine (see Fig. 5 and Table 1 in Supplementary Information). Four coexisting inclusions are normal with respect to Sr, but variably depleted in the most incompatible elements, with one inclusion of ultra-depleted type. The fifth inclusion is Sr-rich.

These observations show that the exotic melt fractions are very small in volume and are therefore readily 'lost' in the bulk magmas by mixing processes during the residence time of single olivine crystals. Similar relationships were also found in mid-ocean-ridge basalt (MORB)^{9,10}; they thus appear to be general features of mantle-derived magmas.

Melt inclusions with significant excess of Sr and Ba have been found also on Iceland¹¹ and mid-ocean ridges¹². They have been interpreted as a result of melting of plagioclase-bearing mantle

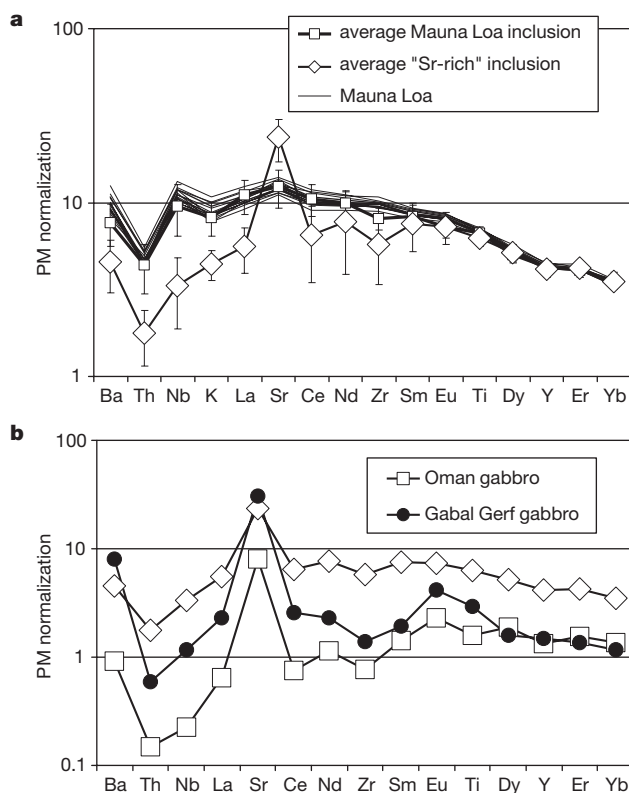


Figure 2 Incompatible-element patterns in melt inclusions, lavas and ophiolitic gabbros normalized to the composition of primitive mantle (PM)³⁰. **a**, Average incompatible-element patterns of normal and Sr-rich melt inclusions in olivine phenocrysts and lavas of Mauna Loa (Hawaii). Error bars represent one standard deviation. The compositions of 15 historical and prehistoric Mauna Loa lavas analysed by isotope dilution mass spectrometry and X-ray fluorescence spectrometry techniques are shown for comparison (A.W.H. *et al.*, unpublished data). Inclusion compositions were reconstructed up to equilibrium with the most Mg-rich olivine (Fo 90.8) indicated in Mauna Loa section of HSDP pilot hole³¹. This leads to 17 ± 0.5 wt% MgO in the melt. Lava compositions were recalculated to 17 wt% MgO using the olivine control line from Fig. 1. **b**, Comparison of incompatible-element patterns in Sr-rich melt inclusions and ophiolitic gabbros. The specific features of Sr-rich inclusions, namely strong Sr positive anomaly, strong negative Th, Nb and Zr anomalies, as well as a moderate positive Ba anomaly, mimic those of gabbros from Oman (C. J. Garrido & P. B. Kelemen, manuscript in preparation) and Gabal Gerf⁶. In gabbros these features are caused by accumulation of plagioclase. The amplitudes of extremes in the incompatible-element patterns of Sr-rich melt inclusions and gabbros are remarkably similar. An exception is the apparent absence of a positive Eu anomaly in melts. This feature is not surprising, partly because a source Sm/Eu ratio will be increased during melting in the presence of garnet (see Supplementary Information) and partly because the expected Eu anomaly (assuming $\text{Eu}^*/\text{Sr}^* = 0.05\text{--}0.1$) is equal to or less than the analytical error of about 20% (see Methods). $\text{Eu}^* = \frac{\text{Eu}_n}{\text{Sm}_n + \text{Gd}_n} - 1$; $\text{Sr}^* = \frac{\text{Sr}_n}{\text{Ce}_n} - 1$.

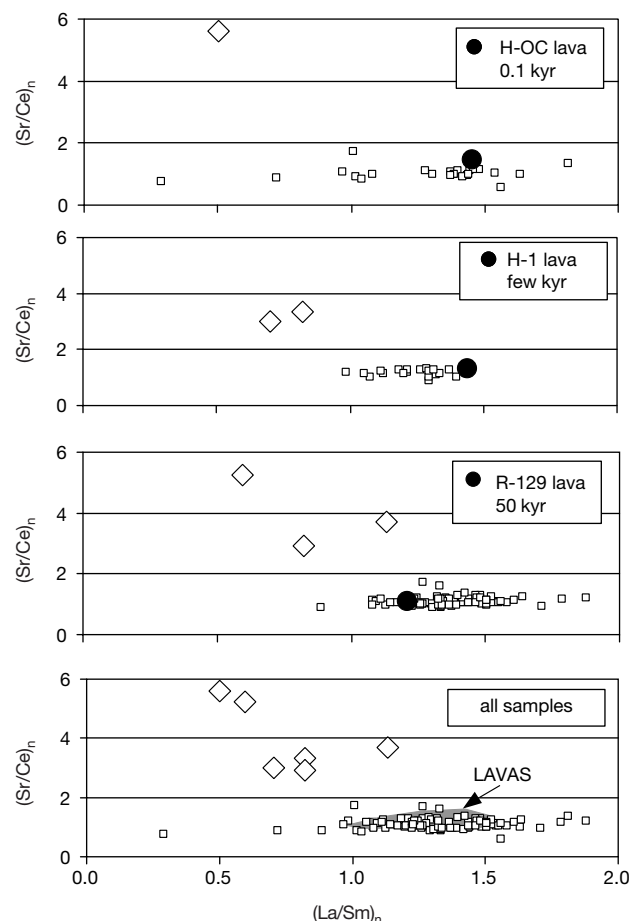


Figure 3 Sr/Ce versus La/Sm ratios of melt inclusions in olivine and host lavas, normalized to primitive mantle values³⁰. Sr-rich melts are present in all three samples. The range of La/Sm ratios of inclusions far exceed those for lavas (ref. 8 and the GEOROC database, see Fig. 1 legend). Furthermore, there is no relationship between the La/Sm ratios of host rocks and melt inclusions: all inclusions of sample H-1 and most inclusions of H-OC are more depleted in La than the host lavas, while the reverse is true for sample R129. This demonstrates that the olivines were not crystallized directly from the melt that transported them at the time of eruption.

sources^{11,12}. This mechanism, irrespective of its merits in the case of Iceland or MORB, is clearly not applicable to melt production beneath Hawaii, which occurs largely, if not completely, in the garnet peridotite facies (see, for example, refs 5, 13).

The incompatible-element patterns of Sr-rich melts are characteristic of oceanic-crust gabbroic rocks enriched in cumulus plagioclase (Fig. 2b). In contrast, the major-element compositions of corrected Sr-rich inclusions are similar to normal inclusions and lavas, and show no enrichment in Al such as would result from assimilation of plagioclase or plagioclase-rich gabbro (Fig. 1). However, the comparison of absolute concentrations of elements in corrected Sr-rich and normal inclusions is subject to the assumption of equal and known initial FeO contents, an assumption which is fundamental to the correction procedure (see Methods). Because the compositions of the Sr-rich melts are unique, this assumption cannot easily be proved for them. Nevertheless, Fig. 4 shows that major-element ratios, such as Al/Ca or Al/Ti, which are not affected either by the correction procedure or by fractional crystallization of olivine, are indistinguishable for Sr-rich and normal melts. At the same time, the relative Sr content varies by up to a factor of 5 between the two groups.

In order to produce such extreme Sr enrichments in the melt by assimilation-type processes or reactive porous flow in the crust, at least 60% of Sr-rich gabbro or 40% of plagioclase would have to be consumed (see Fig. 4 and Supplementary Information). This would drastically affect relative Al concentrations, which is clearly not observed (Figs 1 and 4). This is strong evidence against the production of Sr-rich melts by direct reaction with plagioclase-bearing rocks. (Carbonatites also possess high Sr concentrations, but can be ruled out because, unlike the Sr-rich inclusions, they also have high Ba, light-REE and Th contents¹⁴). We therefore conclude that the anomalous compositions are inherited from the source of the melts and represent recycled gabbro.

The fate of subducted oceanic lithosphere is not well understood. In particular, it has not been clear to what extent former crust can retain its “chemical identity” in a plume source¹⁵. Here we assume that the crustal component has not been completely mixed with the peridotitic matrix but retained a separate identity, thus forming a “marble cake” mantle¹⁶. At high pressures, crustal gabbro transforms to eclogite (garnet–clinopyroxene–SiO₂ ± kyanite ± rutile)¹⁷. Eclogites with similar Sr anomalies have indeed been observed in mantle xenoliths¹⁸.

How does such an eclogite-bearing “marble cake” melt? Experimental studies^{19,20} predict that eclogite bodies containing coesite, will start to melt at higher pressures than the ultramafic matrix in an ascending plume. The resulting (Si-rich) melt will infiltrate the surrounding peridotite, react with it to form pyroxene, and modify the composition of garnet²⁰. Thus, the incompatible-element signature of the recycled crust is transferred to the ascending ultramafic source before the latter starts to melt. Subsequent adiabatic melting of such peridotites in the garnet stability field will then produce melts characterized by a ghost plagioclase signature, while residual garnet still buffers Al and heavy-REE at low abundances (see Supplementary Information). This is exactly what distinguishes the composition of the Sr-rich melts.

We emphasize that, because of the incompatible nature of the ghost plagioclase signature, even small amounts (down to 3 wt%) of partial melt originating in the eclogite could produce large geochemical effects without significant change in mineral compositions and proportions in the metasomatized peridotite (see Supplementary Information). This implies that major-element compositions of melts formed in adjacent metasomatized and normal peridotites will be similar.

In this context we note that a low $\delta^{18}\text{O}$ signature, as observed for example in olivines from Mauna Kea²¹, is not a necessary criterion for former or present gabbro to be involved in magma generation. This is partly because oxygen is a major element and its isotopic

composition will not necessarily change significantly in the process discussed above. Moreover, oceanic gabbros do not universally possess such a signature²². For example, the Wadi Hilti gabbro profile of the Oman ophiolite²², and most gabbros drilled in site 735 in actual oceanic crust²³, are essentially normal in $\delta^{18}\text{O}$. The absence of such a signature in Mauna Loa olivines²¹ can therefore not be construed as evidence against the presence of a gabbroic fingerprint.

Why do normal melt inclusions and lavas have higher incompatible-element concentrations than Sr-rich melts? We suggest that their mantle source is also affected by recycled crust, but this crustal component is closer to average oceanic crust than cumulus-plagioclase-rich, incompatible-element-depleted gabbro. We therefore visualize the plume source as a mixture of peridotite enriched in components of recycled oceanic crust, only one of which is plagioclase-rich gabbro. The accumulation of plagioclase is a result of crystal settling during fractional crystallization. This limits the spatial scale of such heterogeneities to a maximum of a few kilometres or even less, if the data on oceanic gabbros of Hart *et al.*²³ are representative of lower oceanic crust. Mixing these local heterogeneities with the rest of the oceanic crust would erase the ghost plagioclase signature. Thus, our results suggest that such

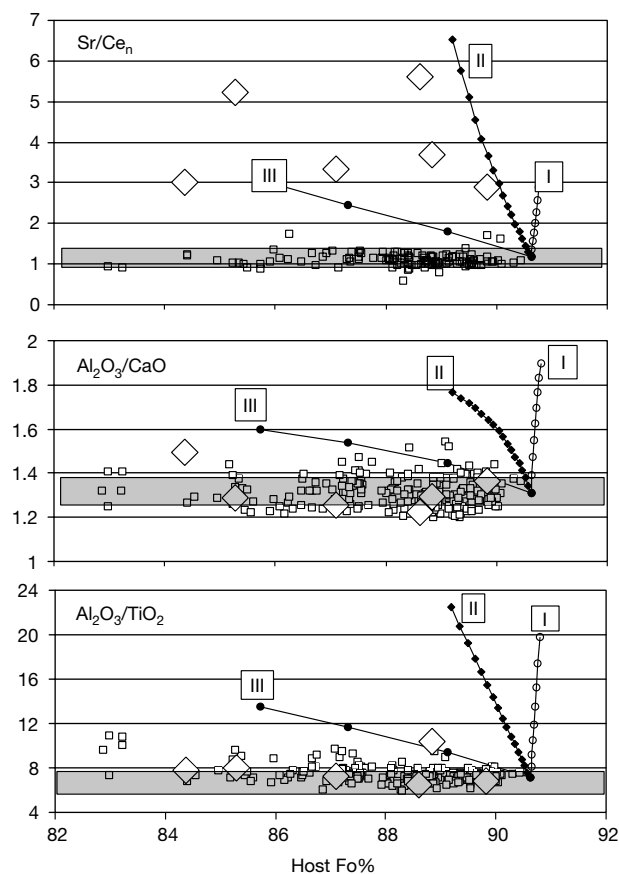


Figure 4 Sr/Ce, Al₂O₃/CaO and Al₂O₃/TiO₂ ratios in melt inclusions versus forsterite content of host olivine. Shaded areas show ranges of lava compositions from GEOROC database (see Fig. 1 legend). Note that Sr-rich melts were trapped by the olivines in nearly the entire range of their composition. The numbered curves show the effects of assimilation of plagioclase from Gabal Gerf gabbro (model I: up to 40 wt% assimilation), assimilation of Gabal Gerf gabbro (model II: up to 80 wt% assimilation) and of assimilation-fractional-crystallization processes (model III: up to 60 wt% assimilated gabbro and 15% olivine fractionation) (see Supplementary Information). The results show that interaction with gabbro sufficient to increase Sr/Ce ratios to levels observed in Sr-rich inclusions would require large and correlated effects on Al/Ca and Al/Ti ratios, none of which is observed in the Sr-rich inclusions.

mixing is incomplete at best, and that some local heterogeneities do survive the entire recycling process.

In contrast with the source heterogeneity required to produce Sr-rich melts, the extreme fractionation between elements with different incompatibility—demonstrated by La/Sm in Fig. 3 for the normal melts—can be produced by the melting process itself. This requires a melting process with efficient extraction of small melt fractions of less than 2 wt% for classical mantle plumes, similar to mid-oceanic ridges⁹.

The extreme heterogeneity of the trace-element patterns requires that some melts preserve the source heterogeneities up to the point where they are incorporated into the olivine phenocrysts. Because the olivines crystallized at low pressure (see Methods), this appears to take place in crustal magma chambers, and therefore small, compositionally distinct magma batches must be transported from mantle to crust, perhaps through separate channels arranged in “fractal trees”²⁴. The much greater homogeneity of the erupted Mauna Loa lavas is the result of subsequent mixing processes²⁵. □

Methods

Samples

The studied historic (1868) H-OC and prehistoric H-1 samples were described in detail elsewhere¹³. Sample R129-8.1 came from 210 m depth (unit 32) of the pilot hole of the Hawaii Scientific Drilling Project (HSDP)²⁶ and has been dated at approximately 50 kyr (ref. 27). Olivines were separated from samples H-OC and H-1 by hand-picking from coarse fragments (0.5–1 mm) of lava samples crushed by hand. Sample R129-8.1 was sequentially sectioned, and olivines were picked directly from doubly polished ‘thick sections’ (200–300 µm thick). This rules out the possibility that olivines with exotic melt inclusions might be introduced by sample contamination.

Inclusions

Melt inclusions in olivine from H-OC and H-1 samples were crystallized as aggregates of clinopyroxene and glass. In order to reconstruct original compositions of trapped melt, these inclusions were heated on an optical heating stage until homogenization and were subsequently quenched¹³.

Most inclusions in olivine from sample R129-8.1 are present as fresh glass with shrinkage bubbles, indicating crystallization of host olivine on the walls of the cavity²⁸. Secondary alteration products are present in cracks but do not affect the main part of the inclusions, which consist of optically clear glass. The Sr-rich melts are not affected by secondary alteration, which is demonstrated by their low K and H₂O contents in addition to the optical observations (see Supplementary Information Table 1). No plagioclase has been observed either as a phase in melt inclusions or as phenocrysts or inclusions in olivine in all studied samples.

The pressure of entrapment of melt inclusions of both normal and exotic types is believed to be around 1 kbar. This conclusion is based on the following arguments: (1) melt inclusions commonly coexist with low-density CO₂-rich fluid inclusions corresponding to entrapment pressures of up to 1.3 kbar and partial H₂O pressures of 5–20 bar (ref. 13). Although there were no fluid inclusions found directly in association with exotic melt inclusions, their H₂O contents (0.2–0.3 wt%) fit the latter pressures. (2) The host olivines for all inclusions correspond to the high-Ca (CaO = 0.2–0.25 wt%) low-pressure variety¹³. (3) Optical heating experiments with exotic melt inclusions (olivine grains H-1-27, H-1-II/70 and HOC-76; see Supplementary Information Table 1) and normal inclusions lead to their homogenization without decrepitation which would be expected if internal pressures exceeded 3–5 kbar in large inclusions (more than 30 µm in diameter) in olivine^{13,28}. (4) Exotic and normal melt inclusions coexist in single olivine grains.

Electron microprobe analysis

Inclusions and host olivine were studied by the Jeol Superprobe electron probe at the Department of Earth Sciences, Mainz University.

Ion microprobe analysis

Inclusions were analysed by ion microprobe ims-4f (Cameca) at the Institute of Microelectronics, Russian Academy of Science (Yaroslavl, Russia) using methods presented elsewhere¹⁰. In order to improve beam stability and avoid mass superposition, we used O²⁻ instead of O⁻ as a primary beam and a 100 kV high voltage offset. The resulting precision, as checked by standards (analysed as unknowns) is better than 10% relative for most elements, and around 20% relative for Eu (due to mass superposition with BaO) and Th where contents are below 0.50 p.p.m. In order to avoid problems caused by boundary layer effects in small inclusions²⁹, only relatively large melt inclusions with minimum dimensions of more than 30 µm were studied.

Reconstruction of composition of inclusions

All natural glassy inclusions in sample R129-8.1 show evidence of quenched olivine crystallization on the cavity walls. An attempt to remove this effect by ‘reversing’ this

crystallization back to equilibrium with the host olivine⁹ leads, however, to unrealistically low FeO in the melts compared with lavas. Moreover, the FeO contents in the glass are positively correlated with inclusion sizes. This effect is caused by the continued exchange of Fe–Mg between included melt and host olivine in response to the cooling of the magmatic system in equilibrium with constant olivine composition²⁹. In order to remove this effect, the correction procedure as described by Danyushevsky *et al.*²⁹ was applied for all inclusions, assuming a uniform initial FeO content of 11.2 wt% for all melts (corresponding to the average value of primitive Mauna Loa lavas; see GEOROC database at <http://georoc.mpch-mainz.gwdg.de>) and equilibrium with host olivine at an oxygen fugacity corresponding to quartz–fayalite–magnetite buffer¹³. Trace-element concentrations in olivine were assumed to be zero.

Received 12 August 1999; accepted 15 March 2000.

- Hofmann, A. W. & White, W. M. Mantle plumes from ancient oceanic crust. *Earth Planet. Sci. Lett.* **57**, 421–436 (1982).
- Lassiter, J. C. & Hauri, E. H. Osmium-isotope variations in Hawaiian lavas: Evidence for recycled oceanic lithosphere in the Hawaiian plume. *Earth Planet. Sci. Lett.* **164**, 483–496 (1998).
- Hauri, E. H. Major element variability in the Hawaiian mantle plume. *Nature* **382**, 415–419 (1996).
- Blichert-Toft, J., Frey, F. A. & Albarede, F. Hf isotope evidence for pelagic sediments in the source of Hawaiian basalts. *Science* **285**, 879–882 (1999).
- Hofmann, A. W. & Jochum, K. P. Source characteristics derived from very incompatible trace elements in Mauna Loa and Mauna Kea basalts, Hawaiian Scientific Drilling Project. *J. Geophys. Res.* **101**, 11831–11839 (1996).
- Zimmer, M., Kröner, A., Jochum, K. P., Reischmann, T. & Todt, W. The Gabal Gerf complex: A Precambrian N-MORB ophiolite in the Nubian Shield, NE Africa. *Chem. Geol.* **123**, 29–51 (1995).
- Danyushevsky, L. V. The effect of small amounts of H₂O on fractionation of mid-ocean ridge magmas. *Eos* **79** (suppl.), S375 (1998).
- Rhodes, J. M. & Hart, S. R. in *Mauna Loa Revealed: Structure, Composition, History And Hazards* (eds Rhodes, J. M. & Lockwood, J. P.) 263–288 (Geophysical Monograph No. 92, American Geophysical Union, Washington, DC, 1995).
- Sobolev, A. V. & Shimizu, N. Ultra-depleted primary melt included in an olivine from the Mid-Atlantic Ridge. *Nature* **363**, 151–154 (1993).
- Sobolev, A. V. Melt inclusions in minerals as a source of principal petrological information. *Petrology* **4**, 209–220 (1996).
- Gurenko, A. A. & Chaussidon, M. Enriched and depleted primitive melts included in olivine from Icelandic tholeiites - origin by continuous melting of a single mantle column. *Geochim. Cosmochim. Acta* **59**, 2905–2917 (1995).
- Kamenetsky, V. S. *et al.* Calcic melt inclusions in primitive olivine at 43 degrees N MAR: evidence for melt-rock reaction/melting involving clinopyroxene-rich lithologies during MORB generation. *Earth Planet. Sci. Lett.* **160**, 115–132 (1998).
- Sobolev, A. V. & Nikogosian, I. K. Petrology of long-lived mantle plume magmatism: Hawaii (Pacific) and Reunion Island (Indian Ocean). *Petrology* **2**, 111–144 (1994).
- Veksler, I. V., Petitbon, C., Jenner, G. A., Dorfman, A. M. & Dingwell, D. B. Trace element partitioning in immiscible silicate-carbonate liquid systems: an initial experimental study using a centrifuge autoclave. *J. Petrol.* **39**, 2095–2104 (1998).
- Christensen, U. R. & Hofmann, A. W. Segregation of subducted oceanic crust in the convecting mantle. *J. Geophys. Res.* **99**, 19867–19884 (1994).
- Allège, C. J. & Turcotte, D. L. Implications of a two-component marble-cake mantle. *Nature* **323**, 123–127 (1986).
- Sobolev, S. V. & Babeyko, A. Y. Modeling of mineralogical composition, density and elastic-wave velocities in anhydrous magmatic rocks. *Surv. Geophys.* **15**, 515–544 (1994).
- Snyder, G. A. *et al.* The origins of Yakutian eclogite xenoliths. *J. Petrol.* **38**, 85–113 (1997).
- Yasuda, A., Fujii, T. & Kurita, K. Melting phase relations of an anhydrous mid-ocean ridge basalt from 3 to 20 GPa: Implications for the behavior of subducted oceanic crust in the mantle. *J. Geophys. Res.* **99**, 9401–9414 (1994).
- Yaxley, G. M. & Green, D. H. Reactions between eclogite and peridotite: mantle refertilisation by subduction of oceanic crust. *Schweiz. Mineral. Petrogr. Mitt.* **78**, 243–255 (1998).
- Eiler, J. M., Farley, K. A., Valley, J. W., Hofmann, A. W. & Stolper, E. M. Oxygen isotope constraints on the sources of Hawaiian volcanism. *Earth Planet. Sci. Lett.* **144**, 453–468 (1996).
- Stakes, D. S. & Taylor, H. P. The northern Samail ophiolite - an oxygen isotope, microprobe, and field-study. *J. Geophys. Res.* **B 97**, 7043–7080 (1992).
- Hart, S. R., Blusztajn, J., Dick, H. J. B., Meyer, P. S. & Muehlenbachs, K. The fingerprint of seawater circulation in a 500-meter section of ocean crust gabbros. *Geochim. Cosmochim. Acta* **63**, 4059–4080 (1999).
- Hart, S. R. Equilibration during mantle melting - a fractal tree model. *Proc. Natl Acad. Sci. USA* **90**, 11914–11918 (1993).
- Rhodes, J. M. Geochemistry of the 1984 Mauna Loa eruption - implications for magma storage and supply. *J. Geophys. Res.* **B 93**, 4453–4466 (1988).
- Stolper, E. M., DePaolo, D. J. & Thomas, D. M. Introduction to special section: Hawaii Scientific Drilling Project. *J. Geophys. Res.* **101**, 11593–11598 (1996).
- Sharp, W. D., Turrin, B. D., Renne, P. R. & Lanphere, M. A. The ⁴⁰Ar/³⁹Ar and K/Ar dating of lavas from the Hilo 1-km core hole, Hawaii Scientific Drilling Project. *J. Geophys. Res.* **101**, 11607–11616 (1996).
- Roedder, E. Fluid inclusions. *Rev. Mineral.* **12**, 1–644 (1984).
- Danyushevsky, L. V., Della-Pasqua, F. N. & Sokolov, S. Re-equilibration of melt inclusions trapped by magnesian olivine phenocrysts from subduction-related magmas: petrological implication. *Contrib. Mineral. Petrol.* **138**, 68–83 (2000).
- Hofmann, A. W. Chemical differentiation of the Earth: the relationship between mantle, continental crust, and oceanic crust. *Earth Planet. Sci. Lett.* **90**, 297–314 (1988).
- Garcia, M. O. Petrography, olivine and glass chemistry of lavas from the Hawaii Scientific Drilling Project. *J. Geophys. Res.* **101**, 11701–11713 (1996).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank the HSDP team for providing samples. We thank S. Simakin for ion probe analyses of inclusions; L.V. Danyushevsky for the access to PETROLOG thermodynamic modelling software; S.V. Sobolev for modelling phase compositions at high T-P; P. Kelemen for providing unpublished data on Oman Gabbro; E. Macsenaere-Riester for help with the electron microprobe analyses; F. K  nstler for preparing doubly polished sections; and F. Frey, J. Eiler, L.V. Danyushevsky, V. S. Kamenetsky, A. A. Gurenko and S. R. Hart for comments that helped to improve the clarity of the manuscript. This work was supported by Deutsche Forschungsgemeinschaft and the Russian Foundation of Basic Research (A.V.S. and I.K.N.) and an Alexander von Humboldt award (A.V.S.).

Correspondence and requests for materials should be addressed to A.V. S. (e-mail: asobolev@mpch-mainz.mpg.de)

Human population in the biodiversity hotspots

Richard P. Cincotta, Jennifer Wisnewski & Robert Engelman

Research Department, Population Action International, 1300 19th Street, NW, 2nd Floor, Washington DC 20036, USA

Biologists have identified 25 areas, called biodiversity hotspots, that are especially rich in endemic species and particularly threatened by human activities. The human population dynamics of these areas, however, are not well quantified. Here we report estimates of key demographic variables for each hotspot, and for three extensive tropical forest areas¹ that are less immediately threatened. We estimate that in 1995 more than 1.1 billion people, nearly 20% of world population, were living within the hotspots, an area covering about 12% of Earth's terrestrial surface. We estimate that the population growth rate in the hotspots (1995–

2000) is $1.8\% \text{ yr}^{-1}$, substantially higher than the population growth rate of the world as a whole ($1.3\% \text{ yr}^{-1}$) and above that of the developing countries ($1.6\% \text{ yr}^{-1}$). These results suggest that substantial human-induced environmental changes are likely to continue in the hotspots and that demographic change remains an important factor in global biodiversity conservation. The results also underline the potential conservation significance of the continuing worldwide declines in human fertility and of policies and programs that influence human migration.

In 1988, ecologist Norman Myers introduced the term 'biodiversity hotspots' to distinguish a global set of high-priority terrestrial ecoregions for conservation². Myers and others argue that, because their 25 hotspots are high in species endemism and low in pristine vegetation (< 30% remaining), wise conservation investments in these ecoregions could help minimize future extinctions^{2–4}. Primatologist Russell Mittermeier subsequently developed a complementary concept, the 'major tropical wilderness areas'⁵. These three areas of tropical forest (Upper Amazonia/Guyana Shield, the Congo Basin, and the New Guinea/Melanesian Islands) are the most pristine of all terrestrial ecoregions exhibiting a high degree of species endemism. Together they cover 6.3% of Earth's terrestrial surface, an area larger than the United States or China. Myers, Mittermeier and others propose a strategy of conservation investments in these areas as a back-up strategy for the hotspot approach¹. By using mapped world distributions of humans (Fig. 1), various census sources and ecoregional boundary data, we calculated population density and growth rates for each of the biodiversity hotspots and major tropical wilderness areas (see Methods).

We estimate that in 1995 population density in the hotspots was 73 people km^{-2} , a figure 71% greater than that of the world as a whole (excluding ice- or rock-covered land). We found that 16 of the 25 hotspots (Fig. 2a) have population densities at or above the world average (42 people km^{-2}). According to our estimates, from 1995 to 2000, human population was still growing in all but one of the hotspots (the Caucasus), with 19 of the hotspot populations

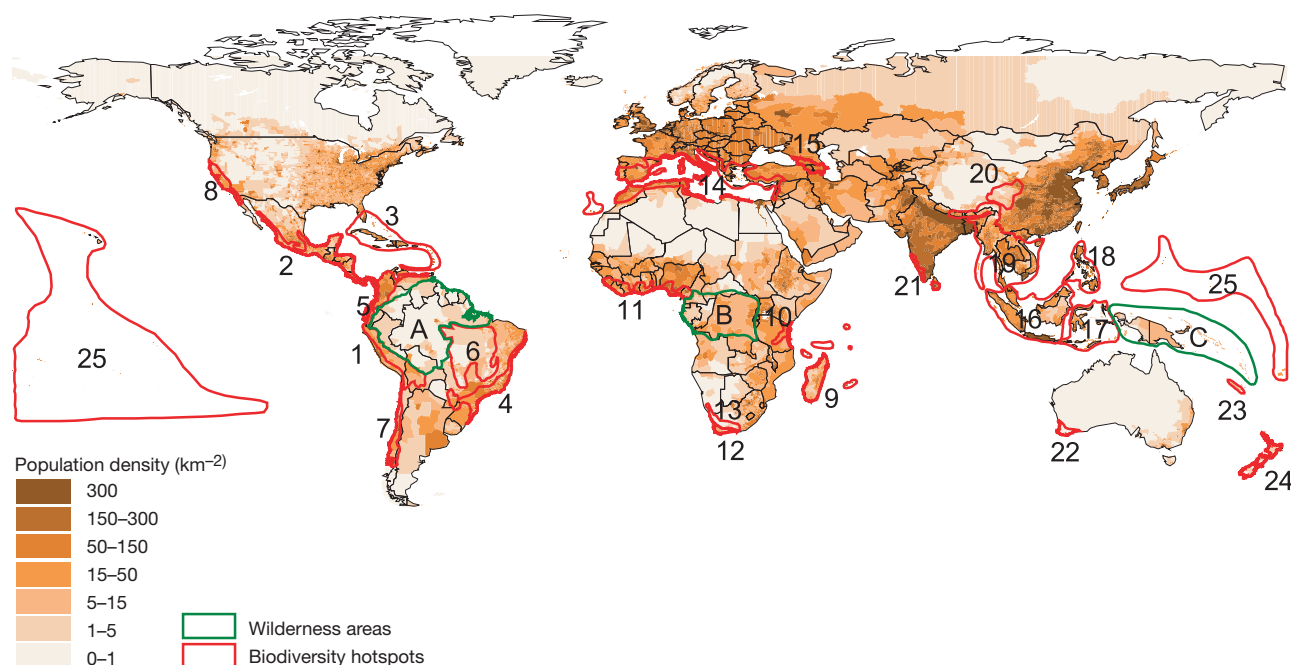


Figure 1 World population density (1995) and the 25 biodiversity hotspots (outlined in red, numbered), and three major tropical wilderness areas (outlined in green, lettered). Hotspots: (1) Tropical Andes; (2) Mesoamerica; (3) Caribbean; (4) Atlantic Forest Region; (5) Choc  -Dari  n-Western Ecuador; (6) Brazilian Cerrado; (7) Central Chile; (8) California Floristic Province; (9) Madagascar; (10) Eastern Arc Mountains and Coastal Forests of Tanzania and Kenya; (11) West African Forests; (12) Cape Floristic Region; (13) Succulent

Karoo; (14) Mediterranean Basin; (15) Caucasus; (16) Sundaland; (17) Wallacea; (18) Philippines; (19) Indo-Burma; (20) Mountains of South-Central China; (21) Western Ghats and Sri Lanka; (22) Southwest Australia; (23) New Caledonia; (24) New Zealand; and (25) Polynesia and Micronesia. Major tropical wilderness areas: (A) Upper Amazonia and Guyana Shield; (B) Congo River Basin; and (C) New Guinea and Melanesian Islands.

growing more rapidly than that of the world as a whole (Fig. 2b). Although population growth rates were, in general, highest in the 19 hotspots wholly within developing countries, growth rates in the hotspots within developed countries were in most cases substantially higher than the worldwide average for developed regions ($0.3\% \text{ yr}^{-1}$).

In 1995, nearly 75 million people (1.3% of world population) were living within the three major tropical wilderness areas, representing an average density of about 8 people km^{-2} . (Area boundaries enclose several major cities.) These areas are experiencing population growth at a rate of $3.1\% \text{ yr}^{-1}$, which is more than twice the global rate.

If population numbers are examined in isolation of other factors, the three hotspots with the most elevated risks, as assessed by high human population density, are the Western Ghats/Sri Lanka, Philippines and Caribbean hotspots. Chocó-Darién-Western Ecuador, Tropical Andes and Madagascar head a list of hotspots facing elevated risks on the basis of rapid population growth alone. Notably, the latest hotspot analysis by Mittermeier *et al.* concludes that the Philippines, Caribbean and Madagascar hotspots appear to be the highest-priority of these ecoregions on the basis of their extreme endemism and the intense packing of species into a much reduced area of original vegetation⁶.

Human population variables are imperfect indicators of risk to biodiversity. Population density figures, for example, obscure

patterns of population distribution within areas. Roughly 90% of the population of the southwest Australia hotspot lives in and around Perth, a single metropolitan area covering less than 2% of the ecoregion; however, such uneven distribution does not negate risk to biodiversity. There is considerable evidence of the capacity of urban populations to alter ecosystems, which are sometimes more than 100 km away, through demand for wood fuel (principally in developing countries), water, food (including wild foods), waste disposal and recreation (mostly in developed countries)^{7,8}.

Another problem is that disturbance caused by humans can occur in the absence of widespread human settlement. This is the frequent result of over-logging, burning, grazing, mining and commercial hunting that have extracted or degraded natural resources, abetted biological invasion or polluted soil and water resources⁹. Population density remains low, for example, in the most arid hotspot, the Succulent Karoo, which experiences heavy grazing and the over-harvesting of its flora for the international trade in ornamental plants.

Population growth rates can also be misleading indicators of risk to biodiversity. Because growth rates are calculated as the annual percentage change in a population, low rates of growth in dense populations add more individuals than much higher rates of growth in sparse populations. Population growth rates mask spatial distributions of growth and the trend of that rate. And both density and growth rates hide the culture, affluence and technology of the numbers of people they represent.

Despite these caveats, however, population trends in the biodiversity hotspots and major tropical wilderness areas indicate a high risk that habitats will continue to degrade as ecosystems dominated by humans expand and species become extinct in the world's most biologically diverse terrestrial regions. Results of the analysis also suggest that, whatever species conservation strategies ultimately emerge, conservation scientists and policymakers should take human population dynamics into account. Especially relevant are trends and potential changes in population growth, density and migration, and the social and economic factors known to influence population variables. One hopeful sign for the conservation of biodiversity is that declines in human fertility are gradually slowing population growth worldwide. □

Methods

Population density

We estimated population densities for biodiversity hotspots and major tropical wilderness areas (ecoregions) using the *Gridded Population of the World, 1995*, a geographic information systems (GIS) layer developed by geographers at the National Center for Geographic Information and Analysis, University of California, Santa Barbara¹⁰. The authors of this layer call attention to numerous sources of potential error in these data, including extrapolation from census-year estimates to 1995 projections, the mapping of census geographical boundaries and census estimates themselves. For census data in most industrialized countries, demographers regularly assume an error (most often an undercount) of less than 3% of the actual population¹¹. Errors exceeding 10% occasionally occur in censuses in the poorest countries, particularly those experiencing political instability. Moreover the populations enumerated vary from country to country, with some countries including, for example, military personnel living outside the country.

Population growth

We partitioned hotspots into countries and sub-national political divisions (provinces or states), used available growth rates or census and projection data to determine the growth of each unit, and then calculated average growth rates for the composite ecoregion. Data on a provincial level were used in ecoregions covering parts of Argentina, Australia, Brazil, Bolivia, Colombia, China, Ecuador, France, India, Indonesia, Mexico, Panama, Peru, South Africa, Spain, Turkey, United States and Venezuela. Where provincial data were unavailable or unnecessary (where the entire country fell within the hotspot), country population growth rates were obtained from estimates generated by the United Nations Population Division¹². These United Nations data are also the source for 1995 world population density and 1995–2000 world population growth rates.

Received 20 October 1999; accepted 17 February 2000.

- Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
- Myers, N. Threatened biotas: hotspots in tropical forests. *Environmentalist* **8**, 178–208 (1988).

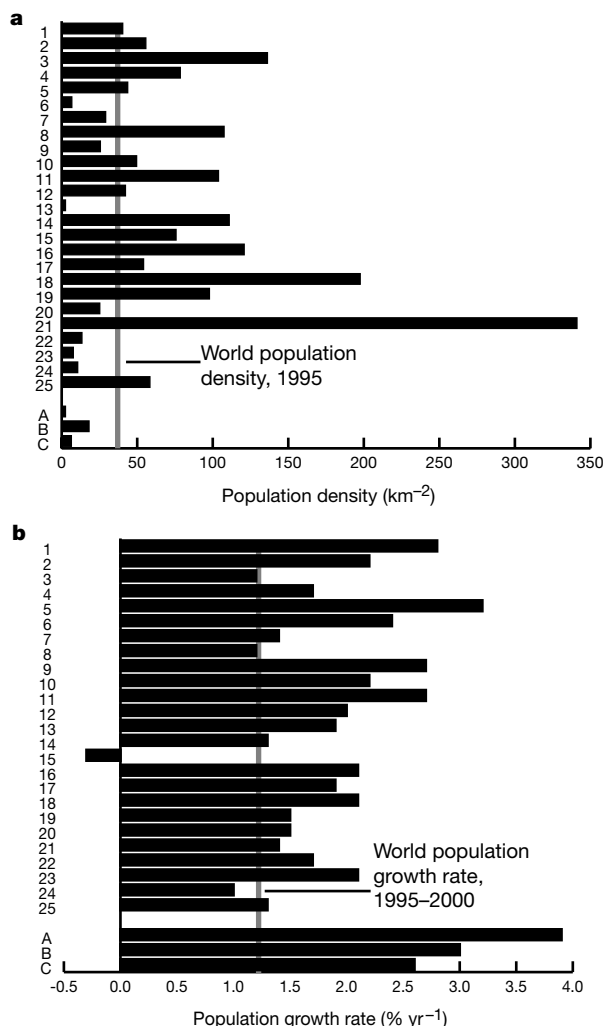


Figure 2 Human population densities (a) and annual growth rates (b) in the 25 global biodiversity hotspots (1–25; see map for names and locations, Fig. 1) and the three major tropical wilderness areas (A, B, C).

3. Myers, N. The biodiversity challenge: expanded hot-spot analysis. *Environmentalist* **10**, 243–256 (1990).
4. Mittermeier, R. A., Myers, N., Thomsen, J. B., da Fonseca, G. A. B. & Olivieri, S. Biodiversity hotspots and major tropical wilderness areas: approaches to setting conservation priorities. *Conserv. Biol.* **12**, 516–520 (1998).
5. Mittermeier, R. A. in *Biodiversity* (eds Wilson, E. O. & Peter, F. M.) 145–154 (National Academy Press, Washington, DC, 1988).
6. Mittermeier, R. A., Myers, N., Robles Gil, P. & Mittermeier, C. G. *Hotspots: Earth's Biologically Richest and Most Threatened Ecosystems* (Cemex, Mexico, D.F., 1999).
7. Repetto, R. The "Second India" Revisited: Population, Poverty, and Environmental Stress Over Two Decades (World Resources Institute, Washington, DC, 1994).
8. Myers, N. in *Population—The Complex Reality* (ed. Graham-Smith, F.) 117–135 (Royal Society, London, 1994).
9. Stedman-Edwards, P. *The Root Causes of Biodiversity Loss: An Analytical Approach* (Worldwide Fund for Nature, Washington, DC, 1997).
10. Tobler, W., Deichmann, U., Gottsegen, J. & Maloy, K. *The Global Demography Project* Tech. Rep. No. 95–6 (National Center for Geographic Information Analysis, Univ. California, Santa Barbara, 1995).
11. Newman, J. L. & Matzke, G. E. *Population: Patterns, Dynamics, and Prospects* (Prentice-Hall, Englewood Cliffs, 1984).
12. U.N. Population Division *World Population Prospects: the 1998 Revision* (United Nations, New York, 1998).

Acknowledgements

We thank A. Bornbusch, D. Blockstein, F. Meyerson, R. Mittermeier, N. Myers and D. Sperling for comments on the manuscript, and K. Sebastian and M. Bartels for solving numerous GIS problems encountered during this research.

Correspondence and requests for materials should be addressed to R.P.C. (e-mail: cincotta@popact.org).

Identification of sleep-promoting neurons *in vitro*

**Thierry Gallopin^{*†‡}, Patrice Fort^{†‡}, Emmanuel Eggermann^{*‡},
Bruno Cauli[§], Pierre-Hervé Luppi[†], Jean Rossier[§], Etienne Audinat[§],
Michel Mühlethaler^{*} & Mauro Serafin^{*}**

^{*} Département de Physiologie, Centre Médical Universitaire, 1 rue Michel-Servet, 1211 Genève 4, Switzerland

[†] Neurobiologie des Etats de Sommeil et d'Eveil, 8 avenue Rockefeller, 69373, Lyon, cedex 08, France

[§] Laboratoire de Neurobiologie et Diversité Cellulaire, CNRS UMR 7637, ESPCI, 10 rue Vauquelin, 75005, Paris, France

[‡] These authors contributed equally to this work

The neurons responsible for the onset of sleep are thought to be located in the preoptic area^{1–3} and more specifically, in the ventrolateral preoptic nucleus (VLPO)^{4–6}. Here we identify sleep-promoting neurons *in vitro* and show that they represent an homogeneous population of cells that must be inhibited by systems of arousal during the waking state. We find that two-thirds of the VLPO neurons are multipolar triangular cells that show a low-threshold spike. This proportion matches that of cells active during sleep in the same region⁶. We then show, using single-cell reverse transcriptase followed by polymerase chain reaction, that these neurons probably contain γ -aminobutyric acid (GABA). We also show that these neurons are inhibited by noradrenaline and acetylcholine, both of which are transmitters of wakefulness^{3,7,8}. As most of these cells are also inhibited by serotonin but unaffected by histamine, their overall inhibition by transmitters of wakefulness is in agreement with their relative inactivity during waking with respect to sleep⁶. We propose that the reciprocal inhibitory interaction of such VLPO neurons with the noradrenergic, serotonergic and cholinergic waking systems to which they project^{5,9,10} is a key factor for promoting sleep.

Intracellular recordings in slices revealed only two cell types within the VLPO. Of 102 recorded cells, most ($n = 70$, 68.6%) were characterized by a potent low-threshold spike (LTS)¹¹ (asterisk and inset in Fig. 1a, LTS cells) that was calcium dependent, as it persisted in tetrodotoxin (TTX, 1 μ M) and was eliminated ($n = 3$) by nickel (200–500 μ M). However, we found no evidence for an intrinsic rhythmicity driven by the LTS¹¹ in these cells. The second, less numerous cell type ($n = 32$, 31.4%) lacked an LTS (Fig. 1b, non-LTS cells) and was usually characterized by a more or less prominent rectification apparent upon depolarization from a hyperpolarized level (Fig. 1b, arrow). Basic membrane parameters, such as resting potential, membrane input resistance and action potential width did not differ between the two cell types. Injection of the intracellular tracer neurobiotin into VLPO neurons indicated that whereas both cell types were medium-sized (LTS cells, $n = 14$; mean large diameter $\pm$ s.d., 19.1 ± 2.0 μ m; mean small diameter, 13.4 ± 1.3 μ m; Fig. 1c, d; non-LTS cells, $n = 6$; 21.3 ± 3.1 μ m versus 11.8 ± 1.3 μ m, respectively; Fig. 1e, f), their shapes and dendritic arbours were completely different. All LTS cells were triangular (Fig. 1d) and multipolar (mean number of primary dendrites: 3.0 ± 0.0 , $n = 14$), whereas non-LTS cells were fusiform (Fig. 1f) and bipolar (1.8 ± 0.4 , $n = 6$).

The high percentage (68%) of LTS cells in the VLPO matches that of cells active during sleep in this region^{4,6} and indicates that the LTS cells may correspond to these sleep-active cells. To test this proposal we measured the effects of noradrenaline, an important transmitter of wakefulness^{3,7,8}, and found that 18 out of 20 LTS cells (Fig. 2a, c) were hyperpolarized by noradrenaline (two were depolarized), whereas all ($n = 8$) non-LTS cells were depolarized (Fig. 2b, c). These results indicate that the LTS cells in the VLPO should be inhibited during waking, when noradrenaline is preferentially released^{3,7,8}, and thus are well suited to correspond to the sleep-active cells recorded *in vivo*^{1,12,13}. Non-LTS cells, in contrast, are not well qualified for that role and will not be considered further here.

The results described above were obtained from intracellular recordings using sharp electrodes. We wanted to test whether VLPO cells are inhibited by noradrenaline in a condition closer to the *in vivo* situation, that is, with minimal perturbation of the cells' properties. We therefore used infrared videomicroscopy¹⁴ to record extracellularly from VLPO triangular multipolar neurons (Fig. 3a) in a loose-attached cell configuration¹⁵. All neurons ($n = 9$) tested in this way were inhibited by noradrenaline (Fig. 3b, c). We then tested whether neurons inhibited by noradrenaline were also inhibited by acetylcholine, another important transmitter of arousal^{3,7,8}; in every case ($n = 5$), these neurons were inhibited by both transmitters (Fig. 3d, e). In addition, the effects of both transmitters were postsynaptic, as they persisted ($n = 2$) in a high magnesium (10 mM)/low calcium (0.1 mM) solution.

We also investigated the two other transmitters (serotonin and histamine) usually associated with arousal^{3,8}. Serotonin (100 μ M, $n = 10$), like acetylcholine, inhibited the majority of cells (7 out of 10) previously inhibited by noradrenaline (Fig. 3f, g) and excited only a minority (3 out of 10). Both effects persisted (respectively, $n = 2$ and $n = 1$) in a high magnesium/low calcium solution. In contrast, histamine (100 μ M, $n = 5$), which was also tested on neurons inhibited by noradrenaline, had no inhibitory or excitatory effect (not shown).

To establish the possible functional role of the LTS cells we needed to identify their neurotransmitter. Most of the VLPO cells, retrogradely labelled from the histaminergic tuberomammillary nucleus⁵, the noradrenergic locus coeruleus⁹ or the cholinergic magnocellular preoptic nucleus¹⁰, are immunoreactive to glutamic acid decarboxylase (GAD) and thus contain GABA. We investigated the expression of GAD in LTS cells using single-cell reverse transcriptase followed by polymerase chain reaction (RT-PCR)^{16–19}. In addition to GAD65 and GAD67, the synthesizing enzymes for GABA, we examined the expression of choline acetyltransferase

(ChAT, the synthesising enzyme for acetylcholine) and enkephalin. Neurons, recorded in the whole-cell mode, were selected in the VLPO on the basis of their multipolar aspect and triangular shape (Fig. 4a, inset) and the presence of an LTS (Fig. 4a). Their cytoplasm was then aspirated into the patch pipette and subsequently analysed by RT-PCR. Whereas two cells (out of eight tested) were found to express enkephalin, all other VLPO neurons (six out of eight) defined as LTS cells expressed GAD65, GAD67 or both (Fig. 4b).

We have identified and described the properties of neurons that have the capacity to promote sleep, and demonstrated their modulation by transmitters of arousal. These cells, located in the VLPO, are triangular and multipolar, show a powerful LTS, contain GABA and are inhibited by noradrenaline and acetylcholine. As most of these cells are also inhibited by serotonin but none are affected by histamine, these results are consistent with their relative inactivity

during waking^{4,6} when all these transmitters are preferentially released³. The complete absence of a histaminergic response in these neurons indicates that, during waking, histamine does not directly modulate them, leaving the inhibitory action of the other transmitters unopposed. At sleep onset, such VLPO neurons could increase their firing⁴⁻⁶ under the influence of circadian inputs (such as from the retina²⁰ and the suprachiasmatic nucleus²¹) and homeostatic factors (such as temperature^{1,6} and sleep-promoting substances^{2,3}). Their increased activity should lead to an inhibition of the monoaminergic and cholinergic nuclei to which they project^{5,9,10}, thereby further increasing their activity through disinhibition. Such a process could facilitate the ability of these neurons to promote sleep and could also have a role during rapid eye movement (REM) sleep, during which VLPO cells are active *in vivo*⁶. It could be proposed that during REM, the deeper inhibition of monoaminergic systems (possibly under the influence

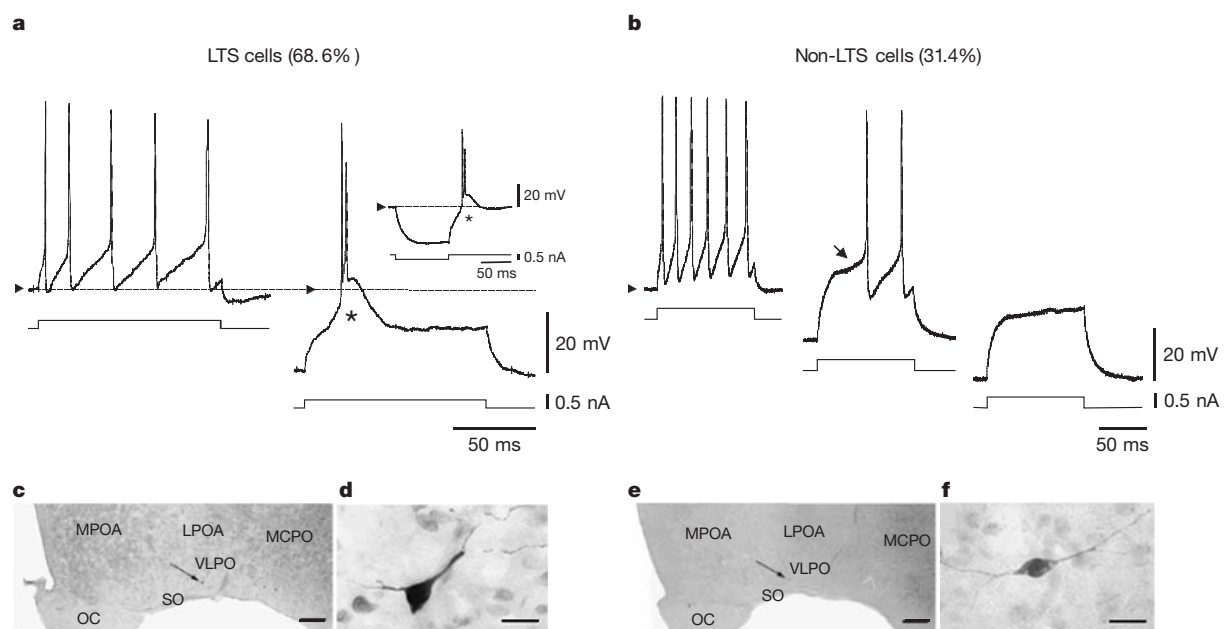


Figure 1 Characterization of VLPO neurons. **a**, LTS cells are characterized by a low-threshold spike (asterisk) when depolarized from a hyperpolarized level. **b**, Non-LTS cells never show low-threshold spikes when they are depolarized from any level of membrane hyperpolarization. They usually show a slowing of the voltage response towards the first action potential (arrow). **c**, Location (arrow) of a neurobiotin-injected LTS cell on a coronal view of the basal forebrain/preoptic area. **d**, Same cell at higher magnification, showing

the characteristic triangular shape of LTS cells. **e**, Location of a non-LTS cell (arrow). **f**, Typical bipolar fusiform aspect of the non-LTS cells. Scale bars, 300 μ m in **c**, **e** and 20 μ m in **d**, **f**. OC, optic chiasm; LPOA, lateral preoptic area; MCPO, magnocellular preoptic nucleus; MPOA, medial preoptic area; VLPO, ventrolateral preoptic nucleus; SO, supraoptic nucleus.

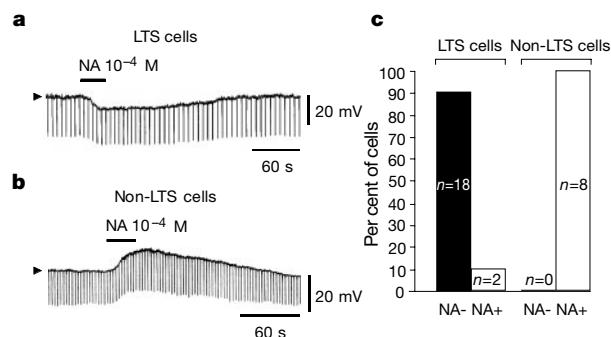


Figure 2 Effects of noradrenaline on VLPO neurons. **a**, Hyperpolarization induced in an LTS cell by a brief (30 s) bath application of noradrenaline (NA). **b**, Depolarization induced by noradrenaline in a non-LTS neuron. **c**, Histogram summarizing data for noradrenaline

(10^{-4} M) application on both types of VLPO cells. Most LTS cells are hyperpolarized by noradrenaline (NA-) and only two are depolarized (NA+). In contrast, all non-LTS cells are depolarized by noradrenaline.

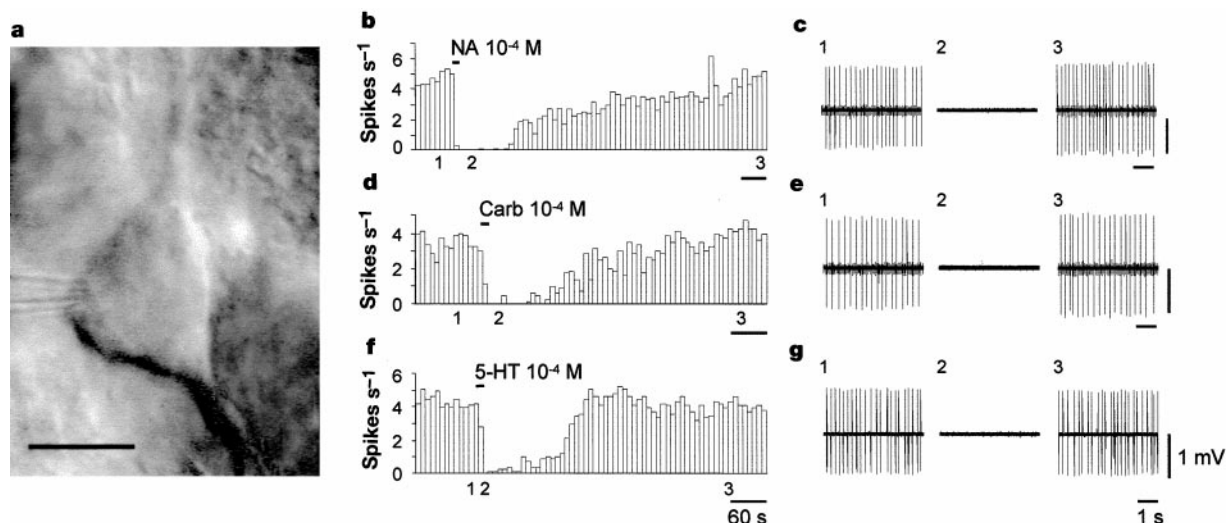


Figure 3 Inhibitory effects of noradrenaline, carbachol and serotonin in a loose-attached patch configuration. **a**, A triangular VLPO neuron with extracellular recording electrode in place. Only two dendrites are visible in the focal plane. Scale bar, 10 μm . **b**, Inhibitory effect of bath-applied (15 s) noradrenaline (NA, 10^{-4} M) on the firing rate of the cell shown in **a**. **c**, Extracellular action potentials taken from positions 1–3 in **b**. **d**, Inhibitory effect of

carbachol (Carb, 10^{-4} M), a non-hydrolysable agonist of acetylcholine, on the cell shown in **a**. **e**, Extracellular action potentials taken from positions 1–3 in **d**. **f**, Inhibitory effect of serotonin (5-HT, 10^{-4} M) on another triangular VLPO neuron previously inhibited by noradrenaline. **g**, Extracellular action potentials taken from positions 1–3 in **f**.

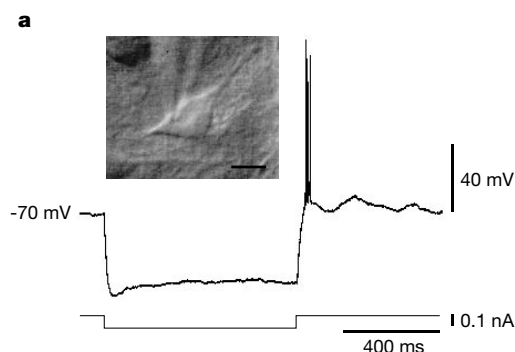


Figure 4 Single-cell RT-PCR on LTS neurons. **a**, Electrophysiological demonstration of an LTS in a VLPO triangular neuron (shown in inset with patch pipette in place; two dendrites are visible in the focal plane; scale bar, 10 μm). **b**, Agarose gel of the PCR

products obtained from the same cell, showing expression of GAD65 and GAD67 and absence of expression of ChAT and enkephalin (Enk). Φ , marker for relative molecular mass (shown to either side).

of ponto–mesencephalic GABA-mediated neurons⁹) might further disinhibit VLPO neurons and enable them to maintain their activity despite the activation of cholinergic neurons during that state³. Finally, it is plausible that cells with characteristics similar to those reported here will be found in other regions of the basal forebrain/preoptic area²², but their aggregation in the VLPO^{4–6}, the direct inputs they might receive from the retina²⁰ and the suprachiasmatic nucleus²¹ and their amenability to *in vitro* investigation, including single-cell RT-PCR, make them ideal candidates for the further study of the preoptic mechanisms involved in sleep generation. □

Methods

Slices and basic electrophysiology

Procedures for preparing rat brain slices have been described²³. Before use, we incubated coronal basal forebrain/preoptic slices (300–400 μm thick) at room temperature in artificial cerebrospinal fluid (ACSF) containing (in mM): NaCl 130, KCl 5, KH_2PO_4 1.25, MgSO_4 1.3, NaHCO_3 20, glucose 10 and CaCl_2 2.4 (0.5 for extracellular recordings), bubbled with a mixture of 95% O_2 and 5% CO_2 . Individual slices were transferred to a thermoregulated (32°C) chamber, under either a dissecting microscope for sharp electrode recordings or a Zeiss Axioskop with an infrared camera¹⁴ for extracellular or whole-cell recordings of identified cells. Slices were maintained immersed and continuously superfused at 3–5 ml min^{-1} with ACSF. For sharp electrode recordings, the pipettes

contained 3 M potassium acetate, whereas for extracellular recordings patch electrodes (5–7 M Ω) were filled with ACSF. We chose sites for recording in rat coronal slices according to the stereotaxic coordinates of the VLPO²⁴ and previously published atlases plotting, at the level of the preoptic area, the VLPO cells that project to the noradrenergic locus coeruleus⁹ and the histaminergic tuberomammillary nucleus⁵. At the lateral confine of the VLPO (in the area corresponding to the magnocellular preoptic nucleus in Fig. 1c, e), we encountered cells with properties similar to those of medial septum/diagonal band cholinergic neurons^{25,26}. Such neurons were considered to be outside the VLPO.

Neurobiotin visualization

After electrophysiological recordings, slices containing one neurobiotin-injected VLPO cell were immersed in an ice-cold fixative with 3% freshly depolymerized paraformaldehyde in 0.1 M phosphate buffer. Neurobiotin-filled neurons were then visualized using the avidin-biotinylated horseradish peroxidase complex reaction (Vectastain, ABC Elite kit, Vector Labs) with 3,3'-diaminobenzidine (Sigma) as a chromogen. To facilitate microscopic observation of the morphology of stained cells and to ensure their location within the VLPO nucleus, slices were re-sectioned at 40 μm using a cryostat. These free-floating sections were then mounted on gelatin-coated slides and counter-stained with 1% Neutral Red to better demarcate the VLPO from the adjacent supraoptic nucleus and magnocellular preoptic nucleus.

Single-cell RT-PCR

Whole-cell recordings were obtained with patch pipettes (3–5 M Ω) filled with 8 μl of internal solution containing (in mM): potassium gluconate 144, MgCl_2 3, EGTA 0.2,

HEPES 10 (pH 7.2; 285–295 mOsm). At the end of the recording, the cell's content was aspirated under visual control into the recording pipette and expelled into a test tube where reverse transcription was performed in a final volume of 10 μ l (ref. 16). We then performed two steps of multiplex PCR. The complementary DNAs present in 10 μ l of the reverse transcription reaction and corresponding to GAD65, GAD67, ChAT and enkephalin were first amplified simultaneously using the primer pairs previously described^{18,19,27}. Taq polymerase (2.5 U, Perkin Elmer-Cetus) and 10 pmol of each of the primers were added to the buffer supplied by the manufacturer (final volume, 100 μ l) and 20 cycles (94 °C, 30 s; 60 °C, 30 s; 72 °C, 35 s) of PCR were run. We then carried out second rounds of PCR using 2 μ l of the first PCR product as template. In this second round, each cDNA was individually amplified using its specific primer pair by performing 35 PCR cycles. We then ran 10 μ l of each individual PCR reaction product on a 1.5% agarose gel using ϕ x174 digested by *Hae*III as markers for relative molecular mass and stained with ethidium bromide. Genomic DNA amplifications, which can occur when the nucleus is harvested, can be easily differentiated from cDNA amplifications by a size criterion. Indeed, for each primer pair, the sense and antisense primers are positioned on two different exons.

Received 29 November 1999; accepted 1 March 2000.

1. Szymusiak, R. Magnocellular nuclei of the basal forebrain: substrates of sleep and arousal regulation. *Sleep* **18**, 478–500 (1995).
2. Shiromani, P. J., Scammell, T., Sherin, J. E. & Saper, C. B. in *Handbook of Behavioral State Control: Cellular and Molecular Mechanisms*. (eds Lydic, R. & Baghdoyan, H. A.) 311–325 (CRC, New York, 1999).
3. Jones, B. E. in *Principles and Practice of Sleep Medicine* (eds Kryger, M. H., Roth, T. & Dement, W. C.) 145–161 (Saunders, Philadelphia, 1994).
4. Sherin, J. E., Shiromani, P. J., McCarley, R. W. & Saper, C. B. Activation of ventrolateral preoptic neurons during sleep. *Neurosci. Res.* **21**, 216–219 (1996).
5. Sherin, J. E., Elmquist, J. K., Torrealba, F. & Saper, C. B. Innervation of histaminergic tuberomammillary neurons by GABAergic and galaninergic neurons in the ventrolateral preoptic nucleus of the rat. *J. Neurosci.* **18**, 4705–4721 (1998).
6. Szymusiak, R., Alam, N., Steininger, T. L. & McGinty, D. Sleep-waking discharge patterns of ventrolateral preoptic/anterior hypothalamic neurons in rats. *Brain Res.* **803**, 178–188 (1998).
7. Steriade, M. & McCarley, R. W. in *Brainstem Control of Wakefulness and Sleep* (eds Steriade, M. & McCarley, R. W.) (Plenum, New York, 1990).
8. McCormick, D. A. & Bal, T. Sleep and arousal: thalamocortical mechanisms. *Annu. Rev. Neurosci.* **20**, 185–215 (1997).
9. Luppi, P.-H. et al. in *Rapid Eye Movement Sleep* (eds Mallett, B. N. & Inoue, S.) 107–122 (Narosa, New Delhi, India, 1999).
10. Fort, P., Gervasoni, D., Rampon, C., Boissard, R. & Luppi, P.-H. GABAergic projections to the magnocellular preoptic area and substantia innominata in the rat. *Neurosci. Abstr.* **24**, 664.3 (1998).
11. Llinás, R. & Yarom, Y. Electrophysiology of mammalian inferior olivary neurones in vitro. Different types of voltage-dependent ionic conductances. *J. Physiol. (Lond.)* **315**, 549–567 (1981).
12. Osaka, T. & Matsumura, H. Noradrenergic inputs to sleep-related neurons in the preoptic area from the locus coeruleus and the ventrolateral medulla in the rat. *Neurosci. Res.* **19**, 39–50 (1994).
13. Osaka, T. & Matsumura, H. Noradrenaline inhibits preoptic sleep-active neurons through α 2-receptors in the rat. *Neurosci. Res.* **21**, 323–330 (1995).
14. Dodd, H. U. & Zieglensberger, W. Infrared videomicroscopy: A new look at neuronal structure and function. *Trends Neurosci.* **17**, 453–458 (1994).
15. Kondo, S. & Marty, A. Synaptic currents at individual connections among stellate cells in rat cerebellar slices. *J. Physiol. (Lond.)* **509**, 221–232 (1998).
16. Lambolez, B., Audinat, E., Bochet, P., Crepel, F. & Rossier, J. AMPA receptor subunits expressed by single Purkinje cells. *Neuron* **9**, 247–258 (1992).
17. Audinat, E., Lambolez, B. & Rossier, J. Functional and molecular analysis of glutamate-gated channels by patch-clamp and RT-PCR at the single cell level. *Neurochem. Int.* **28**, 119–136 (1996).
18. Cauli, B. et al. Molecular and physiological diversity of cortical nonpyramidal cells. *J. Neurosci.* **17**, 3894–3906 (1997).
19. Porter, J. T. et al. Selective excitation of subtypes of neocortical interneurons by nicotinic receptors. *J. Neurosci.* **19**, 5228–5235 (1999).
20. Lu, J., Shiromani, P. & Saper, C. B. Retinal input to the sleep-active ventrolateral preoptic nucleus in the rat. *Neuroscience* **93**, 209–214 (1999).
21. Gaus, S. E. & Saper, C. B. Efferent connections from the suprachiasmatic nucleus to the ventrolateral preoptic nucleus in the rat. *Neurosci. Abstr.* **24**, 762.13 (1998).
22. Fort, P., Khatib, A., Serafini, M., Mühlenthaler, M. & Jones, B. E. Pharmacological characterization and differentiation of non-cholinergic nucleus basalis neurons in vitro. *NeuroReport* **9**, 61–65 (1998).
23. Llinás, R. & Sugimori, M. Electrophysiological properties of in vitro Purkinje cell somata in mammalian cerebellar slices. *J. Physiol. (Lond.)* **305**, 171–195 (1980).
24. Paxinos, G. & Watson, C. *The Rat Brain in Stereotaxic Coordinates* (Academic, San Diego, 1997).
25. Griffith, W. H. Membrane properties of cell types within guinea pig basal forebrain nuclei in vitro. *J. Neurophysiol.* **59**, 1590–1612 (1988).
26. Gorelova, N. & Reiner, P. B. Role of the afterhyperpolarization in control of discharge properties of septal cholinergic neurons in vitro. *J. Neurophysiol.* **75**, 695–706 (1996).
27. Porter, J. T. et al. Properties of bipolar VIPergic interneurons and their excitation by pyramidal neurons in the rat neocortex. *Eur. J. Neurosci.* **10**, 3617–3628 (1998).

Acknowledgements

We thank L. Bernheim, N. Demareux, J. J. Dreifuss and D. Muller for helpful comments on the manuscript and D. Machard for technical assistance. This study was supported by grants from the Swiss Fonds National to M.M. and a French MENRT fellowship to T.G.

Correspondence and requests for materials should be addressed to M.M. (e-mail: michel.muhlethaler@medecine.unige.ch).

TACI and BCMA are receptors for a TNF homologue implicated in B-cell autoimmune disease

Jane A. Gross*, Janet Johnston*, Sherri Mudri*, Rachel Enselman*, Stacey R. Dillon*, Karen Madden†, Wenfeng Xu†, Julia Parrish-Novak†, Don Foster‡, Cathy Lofton-Day§, Margaret Moore§, Alisa Littau‡, Angelika Grossman‡, Harald Haugen‡, Kevin Foley§, Hal Blumberg§, Kim Harrison*, Wayne Kindsvogel* & Christopher H. Clegg*

Departments of *Immunology, †Functional Cloning, ‡In Vivo Biology and §Genetics, ZymoGenetics, 1201 Eastlake Avenue East, Seattle, Washington 98102, USA

B cells are important in the development of autoimmune disorders by mechanisms involving dysregulated polyclonal B-cell activation, production of pathogenic antibodies, and co-stimulation of autoreactive T cells. zTNF4 (BLyS, BAFF, TALL-1, THANK)^{1–5} is a member of the tumour necrosis factor (TNF) ligand family that is a potent co-activator of B cells *in vitro* and *in vivo*^{1,2,5}. Here we identify two receptors for zTNF4 and demonstrate a relationship between zTNF4 and autoimmune disease. Transgenic animals overexpressing zTNF4 in lymphoid cells develop symptoms characteristic of systemic lupus erythematosus (SLE) and expand a rare population of splenic B-1a lymphocytes. In addition, circulating zTNF4 is more abundant in NZBWF1 and MRL-*lpr/lpr* mice during the onset and progression of SLE. We have identified two TNF receptor family members, TACI⁶ and BCMA^{7,8}, that bind zTNF4. Treatment of NZBWF1 mice with soluble TACI-Ig fusion protein inhibits the development of proteinuria and prolongs survival of the animals. These findings demonstrate the involvement of zTNF4 and its receptors in the development of SLE and identify TACI-Ig as a promising treatment of autoimmune disease in humans.

zTNF4 was identified in a human-granulocyte-derived complementary DNA library by homology with other known TNF ligands⁹. Soluble recombinant zTNF4 stimulates proliferation of human B cells in synergy with other B-cell activators, augments immunoglobulin production, and upregulates expression of cell-surface molecules involved in B-cell effector function (refs 1, 2 and data not shown). Recent studies demonstrate that expression of BAFF, in the liver of transgenic mice, results in lymphoid disorders and autoimmune manifestations⁵. We obtained similar results expressing zTNF4 in transgenic mice using a lymphoid specific V_H promoter¹⁰ and E μ enhancer¹¹ (zTNF4-TG). We identified expression of the transgene in 15 founder animals by polymerase chain reaction with reverse transcription (RT-PCR) and a zTNF4 enzyme-linked immunosorbent assay (ELISA) detected increased levels of circulating zTNF4 protein in these animals (data not shown). Flow cytometric analysis revealed a marked increase in the proportion and total number of B220⁺ B cells in the spleen and lymph node relative to controls (Fig. 1a) and an increased percentage of syndecan⁺ plasma cells (Fig. 1b). However, there was no apparent effect of zTNF4 overexpression on development of B220⁺ IgM⁺ progenitor B cells in bone marrow (data not shown). The total number of splenic T cells in the transgenic animals was normal, but the CD4⁺ and CD8⁺ T-cell populations displayed an activated phenotype defined by decreased levels of LECAM-1 and increased CD44 (data not shown).

We analysed the levels of immunoglobulin collected from 15 founders and 9 offspring ranging from 6 to 23 weeks of age, and determined that the amount of both IgM and IgG was elevated at least threefold in over 50% of the animals tested (IgM increased

more than IgG). IgE levels were also increased in 18% of the serum samples from the zTNF4-TG mice (Fig. 1c). Significantly, several older zTNF4-TG animals developed high titres (1:1,600) of anti-double-stranded (ds) DNA antibodies (Fig. 1c), proteinuria (data not shown) and glomerulonephritis (Fig. 1d), suggesting that unregulated expression of zTNF4 can contribute to the development of SLE-like symptoms.

Thus, the phenotype of zTNF4-TG mice is similar to that of BAFF transgenic mice⁵. Unlike BAFF transgenic mice, zTNF4-TG mice also contained elevated numbers of splenic B-1a cells (CD5⁺, B220^{lo} and IgD^{lo}) (Fig. 1b). This B-cell lineage is thought to produce predominantly self-reactive IgM antibodies and represents only a small fraction of splenocytes (0.5%) in normal mice¹². We also identified a population of CD19⁺ B cells (10–15%) in thymuses

from 3 out of 15 zTNF4 founders (Fig. 1b) that may represent expansion of resident B cells or recirculating peripheral B cells. B cells were found to accumulate in other tissues including the lung and small intestine (data not shown).

To characterize the association of zTNF4 with SLE further, we measured zTNF4 protein levels in serum from NZBWF1 and MRL-*lpr/lpr* mice. Both mouse strains develop chronic, spontaneous autoimmune disease¹³. As a baseline, serum zTNF4 was measured in female NZBWF1 animals before onset of disease (10 weeks of age) and then measured during disease progression (Fig. 2a). An average threefold increase in serum zTNF4 was detected in 20% of the animals at 14 weeks of age, and reached sixfold over baseline in 100% of the animals by 24 weeks of age. zTNF4 levels were highest (ninefold over baseline) in animals with

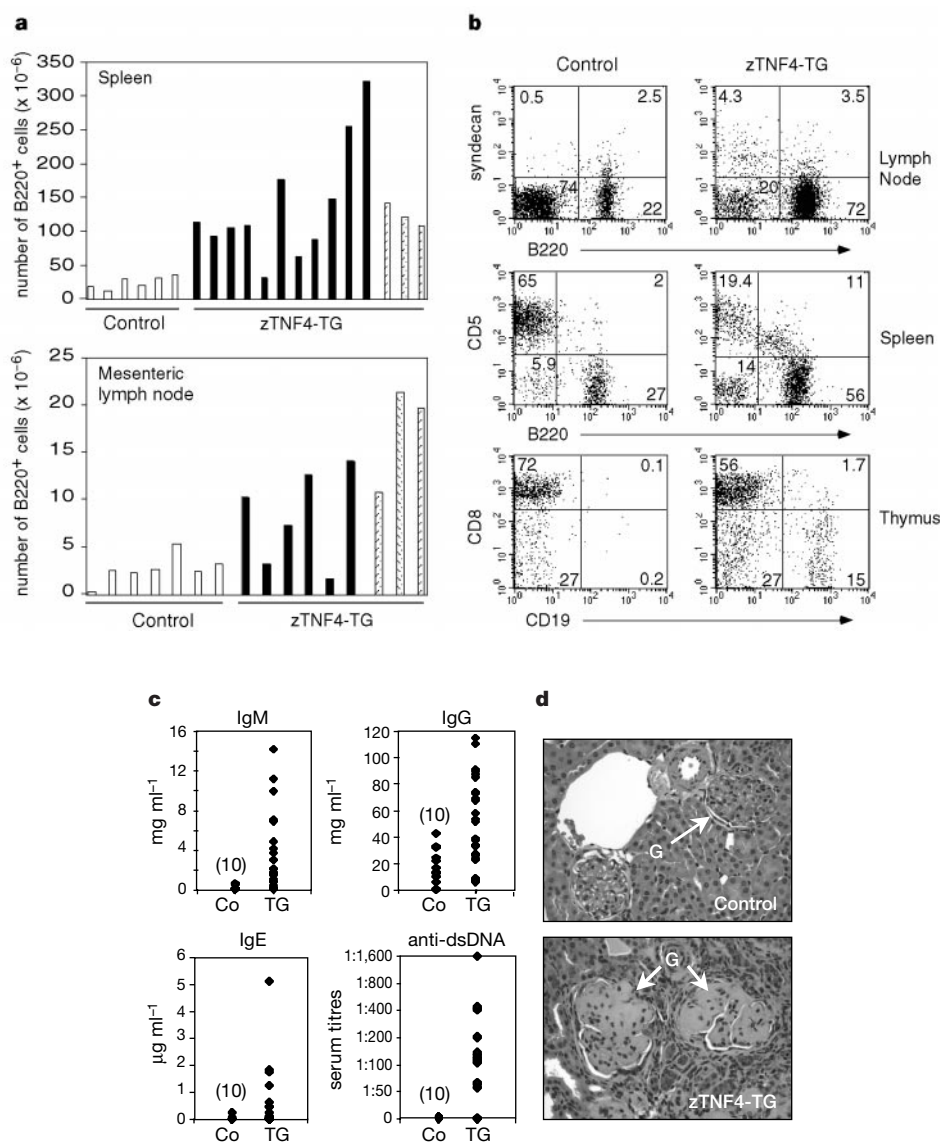


Figure 1 zTNF4 transgenic animals develop a phenotype characteristic of SLE. **a**, The total number of B220⁺ lymphocytes was determined by fluorescence-activated cell sorting (FACS) in cell suspensions prepared from spleen and mesenteric lymph nodes. Each bar represents data from individual control mice (open bars), 13-week-old zTNF4 transgenic (zTNF4-TG) founder animals (shaded bars) and 11-week-old transgenic offspring (hatched bars). **b**, Cells were isolated from lymph node (top), spleen (middle), and thymus (bottom) from zTNF4-TG or control mice and analysed by FACS with the indicated antibodies. Data are representative of 11 individual zTNF4-TG founders, 3

offspring and 6 control mice analysed. **c**, IgM, IgG and IgE and anti-dsDNA titres were assayed by ELISA in 23 serum samples from 14 zTNF4-TG founder animals ranging from 6 to 23 weeks of age, and from 12 age-matched littermate control animals (Co). **d**, Kidney sections from 15-week-old control (top) or zTNF4-TG (bottom) were stained with haematoxylin and eosin. Amyloid deposition and thickened mesangium of the glomeruli (G) were identified. A representative of 4 zTNF4-TG founders and of 12 control mice is shown.

markedly increased titres of anti-dsDNA antibodies and elevated proteinuria ($\geq 2,000 \text{ mg dl}^{-1}$). In contrast, serum zTNF4 was not elevated in the healthy NZB/B parental strain, even at 30 weeks of age. MRL-*lpr/lpr* mice are a more severe model of SLE than the NZBWF1 strain, and rapidly develop autoimmunity and profound lymphoproliferative disease¹⁴. Serum zTNF4 was increased in diseased MRL-*lpr/lpr* mice compared with the amount detected in healthy 11-week-old MRL/MpJ mice (Fig. 2b). However, the relative increase in zTNF4 was similar in both NZBWF1 and MRL-*lpr/lpr* strains when comparing healthy controls with diseased animals.

The receptor for zTNF4 appears to be expressed predominantly on B cells and on activated T cells⁶. Biotinylated zTNF4 was found to bind to almost all mature CD19⁺ peripheral B cells, weakly to immature B cells in the bone marrow, and to most transformed B-cell lines, including RPMI 1788, Raji and Ramos (data not shown). A cDNA expression library, prepared from RPMI 1788 messenger RNA, was transfected into COS7 cells and receptor-positive cells were identified with labelled zTNF4. Two orphan TNF receptor family members, TACI⁶ and BCMA^{7,15}, were isolated. TACI is expressed on B cells and signals through CAML, activating the transcription factors NF-AT, AP-1 and NF κ B⁶. BCMA is a B-cell maturation antigen that is thought to be an intracellular protein^{7,15}. Alignment of TACI and BCMA sequences reveals significant homology between the two receptors within the prototypic cysteine-rich domains⁸ (Fig. 3a), but not with other members of the TNF receptor family. zTNF4 binding to other TNF receptor family members was tested using surface-enhanced laser desorption and ionization

(SELDI) technology. zTNF4 failed to bind 11 out of the 19 known human TNF receptor family members including CD40, p55 TNFR1, p75 TNFR2, Fas, TRAIL R1, R2, R3, R4, OPG and HVEM (data not shown).

Specific binding experiments of ¹²⁵I-labelled zTNF4 were performed on baby hamster kidney (BHK) transfectants expressing either TACI or BCMA. Scatchard plot analysis revealed that TACI and BCMA bound labelled ligand with similar dissociation constants (K_d): 1.25 nM for TACI and 1.11 nM for BCMA (Fig. 3b). This result indicates that BCMA can be expressed on the cell surface under these transfection conditions. Soluble versions of TACI and BCMA were made by fusing their extracellular domains to the Fc portion of human IgG1 (TACI-Ig and BCMA-Ig). These fusion

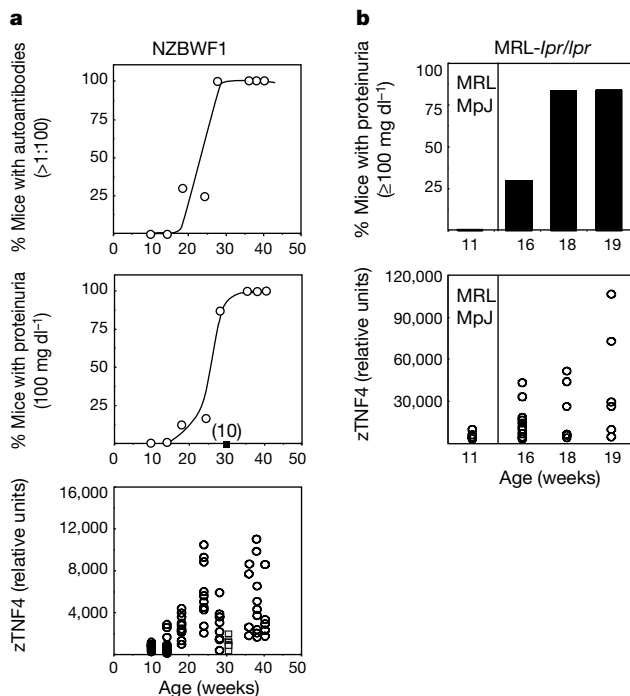


Figure 2 Amount of zTNF4 protein is increased in serum from diseased NZBWF1 mice and MRL-*lpr/lpr* mice. **a**, Autoantibody titres, proteinuria and zTNF4 levels were measured in 68 NZBWF1 mice ranging from 10 to 40 weeks old (open circles), and ten 30-week-old NZB/B control mice (filled squares). Percentages of animals with anti-dsDNA antibody titre $\geq 1:100$ (top panel) and proteinuria $\geq 100 \text{ mg dl}^{-1}$ (middle panel) are shown (each circle and square represents an average of five to ten mice). Amount of zTNF4 protein was measured in a capture assay and is presented as relative zTNF4 units for each individual NZBWF1 mouse (open circles) and control NZB/B mouse (open squares). **b**, Proteinuria (upper panel) and amount of zTNF4 (lower panel) were measured in serum from 23 MRL-*lpr/lpr* mice at ages 16, 18 and 19 weeks and from 10 MRL/MpJ control mice at 11 weeks. Approximate amounts of zTNF4 protein in NZBWF1 (0.1–2 ng ml⁻¹) and MRL-*lpr/lpr* (1–10 ng ml⁻¹) serum were determined using a human zTNF4 standard.

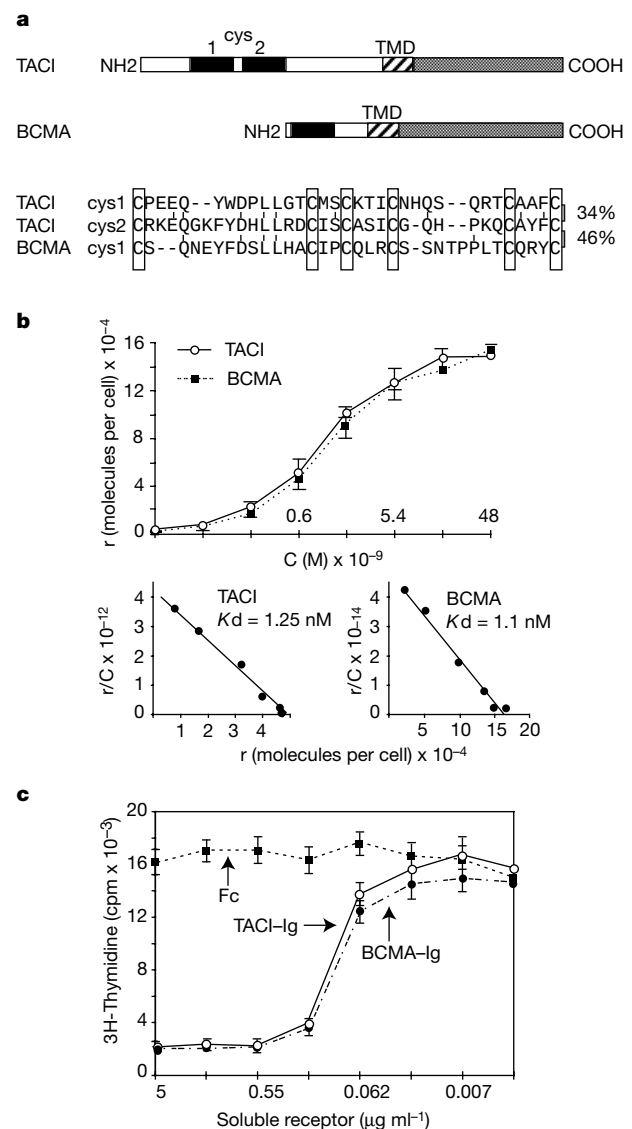


Figure 3 TACI and BCMA are members of the TNF receptor superfamily that bind zTNF4 and can block zTNF4 activity *in vitro*. **a**, TACI and BCMA. The cysteine-rich repeats (cys, black), transmembrane domain (TMD, hatched) and cytoplasmic domains (shaded) are shown. The cysteine-rich repeats from TACI and BCMA are aligned and the percentage of amino-acid identity is indicated. **b**, Binding curves and Scatchard plot analysis of ¹²⁵I-labelled zTNF4 binding to TACI and BCMA were performed using stable BHK transfectants expressing each receptor. **c**, Human peripheral blood B cells were cultured with zTNF4 (25 ng ml⁻¹), IL-4 (10 ng ml⁻¹) and plate-bound anti-IgM (5 μ g ml⁻¹), and treated with increasing amounts of TACI-Ig, BCMA-Ig or control Fc. Cells were cultured for five days and then assayed for cell proliferation by ³H-thymidine incorporation.

proteins prevented zTNF4 binding to human B cells (data not shown) and completely inhibited zTNF4 stimulatory activity on human B cells *in vitro* (Fig 3c) and on murine B cells (data not shown).

To determine the effects of TACI-Ig on the progression of SLE, we treated 21-week-old female NZBWF1 mice (15 mice per group) three times a week for five weeks, with PBS alone, or with 100 µg or 20 µg doses of TACI-Ig or Fc control protein. Treatment with human TACI-Ig (100 µg dose group) delayed the frequency of proteinuria as defined by a marked reduction in the fraction of animals with proteinuria ≥ 100 mg dl⁻¹ for up to 10 weeks after the last treatment ($P < 0.01$ by χ^2 -test) (Fig. 4a). There was also a noticeable but less prominent effect of TACI-Ig at 20 µg per dose on the development of proteinuria. TACI-Ig treatment (100 µg per dose) increased the survival of animals, with 100% surviving at 38 weeks of age (12 weeks after the last treatment) compared with 47% survival in the equivalent Fc treatment group (Fig. 4b). Examination of anti-dsDNA titres revealed no significant difference in autoantibody levels between the different groups (data not shown). Although elevated serum levels of antibodies to nuclear constituents is a hallmark of SLE¹⁶, there is clear evidence that supports the role of B cells in disease progression in the absence of secreted immunoglobulins¹⁷. Alternatively, a longer course of treatment with murine TACI-Ig may be needed to suppress autoantibody titres.

Several mechanisms could contribute to the suppression of disease in the TACI-Ig treated animals. To identify potential alterations in the lymphoid compartment of the TACI-Ig treated SLE mice, we determined the percentage of peripheral blood B cells

(CD5⁻, B220⁺), T cells (CD5⁺, B220⁻) and monocytes (CD11b⁺, B220⁻) by flow cytometry at 28, 31 and 37 weeks of age (Fig. 4c). There was a significant decrease ($P < 0.01$ by Student's *t*-test) in the percentage of peripheral blood B cells at 28 weeks of age in animals treated with 100 µg per dose TACI-Ig (decrease of 53%) and 20 µg per dose TACI-Ig (decrease of 32%) compared with percentages of B cells in dose-matched Fc controls. This reduction in B cells persisted five weeks after the last treatment (31 weeks of age) in the high dose TACI-Ig treatment group and returned to levels found in the control groups at 37 weeks of age. These results indicate a role for zTNF4 in maintaining peripheral B-cell populations that have a pathogenic effect on the development of renal disease.

B cells play a significant role in the development of autoimmunity, through B-cell activation, production of pathogenic antibodies and co-stimulation of autoreactive T cells. zTNF4 is a potent molecule, capable of stimulating B cells and driving B-cell differentiation to a pathogenic state. Understanding the regulation of zTNF4 expression and the biology of its receptors, TACI and BCMA, will be important for understanding B-cell development in the normal animal and during disease states. □

Methods

Recombinant proteins

To produce soluble zTNF4, a DNA fragment containing the sequences for the yeast alpha factor pre-pro leader, a Flag epitope, and the extracellular domain (residue 141–285) of zTNF4 was cloned downstream of the AUG1 promoter into a modified form of the *Pichia methanolica* expression vector pCZR134 (ref. 18). Protein was purified from yeast fermenters on an anti-Flag affinity column. DNA encoding the extracellular domain of human TACI (residue 1–154, ref. 6) or BCMA (residue 1–48, ref. 7) was fused to the Fc region of human Ig heavy chain $\gamma 1$ by PCR and cloned into a mammalian expression vector using the CMV promoter and the TPA pre-propeptide sequence. Proteins were purified by Protein A chromatography from supernatants of transfected CHO DG44 cells adapted into protein-free media (Life Technology Inc.).

Cell culture

zTNF4 stimulations were performed using human B cells purified from donor peripheral blood mononuclear cells by depletion of CD43⁺ cells using anti-CD43 magnetic beads (Miltenyi). Cells (1×10^5) were cultured in RPMI 1640 medium, 10% FCS and L-glutamine in round bottom 96-well plates (Corning) with zTNF4, anti-IgM (Southern Biotechnology) and IL-4 (Pharmingen) as described in the figure legends. Ramos, Raji and RPMI 1788 were purchased from ATCC.

Transgenic mice

Microinjections were performed by standard methodology on B6C3F2 fertilized ova with a construct containing the IgVH promoter and IgE μ enhancer spliced upstream of the SV40 16S intron followed by the zTNF4 open reading frame and a human growth hormone polyadenylation signal sequence^{10,11,19}. Transgenic animals were backcrossed to C57BL/6 mice (JAX). Transgene expression was measured in spleen and thymus RNA samples prepared using RNeasy mini-kit (Qiagen). Samples of total RNA (25 ng) were analysed in duplicate by quantitative RT-PCR using oligonucleotides against the human growth hormone 3' untranslated region. A Gene Amp 5700 sequence detector (Perkin-Elmer) was used to quantitate the PCR reactions compared with a standard curve containing hGH RNA. Single-cell suspensions from transgenic lymphoid tissues were prepared and analysed on a FACSCalibur flow cytometer (Becton Dickinson) using saturating amounts of the appropriate FITC-, PE- and/or TriColour (TC)-conjugated monoclonal antibodies (Pharmingen).

Animal studies

Female NZBWF1 mice were purchased from JAX. The animals were monitored for development of proteinuria every two weeks and serum was collected monthly to measure anti-dsDNA antibody titres. Morbidity was checked three times a week and daily after proteinuria levels reached 2,000 mg dl⁻¹.

Antibody ELISAs and proteinuria

Murine serum immunoglobulins were quantitated by standard ELISA techniques using capture antibodies goat anti-IgG (Kirkegaard and Perry), goat anti-IgM (Zymed) or goat anti-IgE (Pharmingen). Captured mouse Ig was detected using horseradish peroxidase (HRP)-labelled goat anti-IgG, anti-IgM (Jackson) and anti-IgE (Pharmingen). Anti-dsDNA antibodies were measured²⁰ using plates coated with poly(dAdT) (Sigma) detected with HRP-labelled goat anti-mouse IgG Fc (Cappel). The presence of proteins in mouse urine was measured using Uristix (Ames).

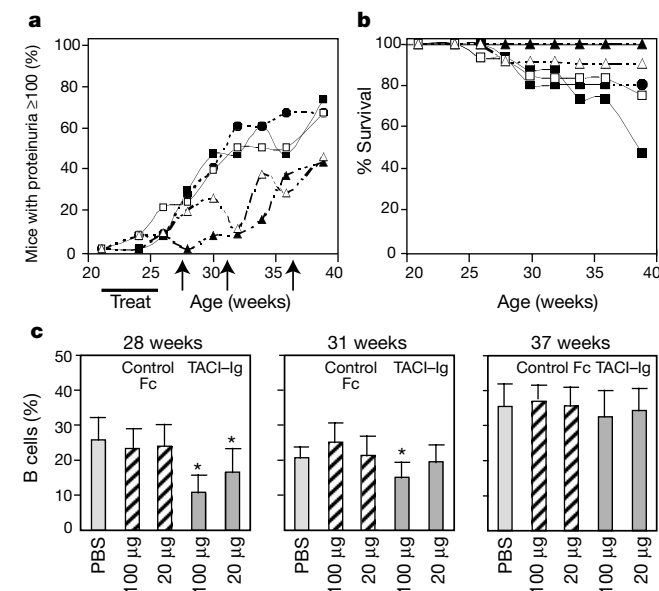


Figure 4 TACI-Ig inhibits the progression of proteinuria and increases the survival of NZBWF1 mice. **a**, Female NZBWF1 mice (15 mice per group) at 21 weeks of age were injected intraperitoneally three times a week for five weeks with PBS (filled circles), control Fc protein at 100 µg per dose (filled squares) or 20 µg per dose (open squares), or TACI-Ig at 100 µg per dose (filled triangles) or 20 µg per dose (open triangles). Proteinuria was assessed every 2 weeks and the mice were bled at 28, 31 and 37 weeks of age (arrows) to determine the percentage of peripheral blood B cells. The percentage of mice in each group with proteinuria ≥ 100 mg dl⁻¹ is shown. **b**, The percentage of mice surviving in each treatment group is plotted as a function of their age. **c**, The percentages of viable B220⁺CD5⁻ B cells in the peripheral blood of mice in each group were determined by flow cytometry at the time points indicated; standard deviations for each group are shown. Statistically significant differences (Student's *t*-test: $P < 0.01$) between TACI-Ig treatment and dose-matched Fc control groups are identified (asterisk).

zTNF4 detection

zTNF4 was detected in a solution-phase capture assay developed for the Origen 1.5 analyser (IGEN Inc.). Briefly, $1 \mu\text{g ml}^{-1}$ of a rabbit anti-zTNF4 polyclonal antibody that was affinity purified and biotinylated was incubated for two hours at 20°C , with undiluted mouse serum samples or a zTNF4 standard curve diluted into normal mouse serum, and $1 \mu\text{g ml}^{-1}$ of a ruthenylated conjugate of the antibody. Streptavidin labelled beads were added at a final concentration of 0.1 mg ml^{-1} and incubated for 30 min at room temperature. The samples were analysed on the Origen 1.5 analyser and reported as electro-chemiluminescence (ECL) units.

zTNF4 receptor cloning

cDNA was prepared from RPMI 1788 cells (ATCC) and cloned into an expression vector (pZP7). This library of ten million cDNAs was transfected into COS7 cells. After 24 h, the cells were fixed²¹ and probed with zTNF4-biotin, followed by a streptavidin-HRP conjugate, and then incubated with an HRP/tyramide substrate to identify receptor-positive cells. Individual positive cells were recovered by micro-dissection. Plasmid DNA was rescued from detergent extracts and used to transform *Escherichia coli*. Individual plasmids were tested by re-transfecting COS7 cells and probing with the zTNF4 probe.

Binding assays

Human zTNF4s protein was iodinated with ^{125}I -Na (Amersham) using Iodo-beads (Pierce) and purified on a Sephadex G25 PD-10 column (Pharmacia). BHK transfectants expressing either TACI or BCMA (2×10^5 cells) were incubated for two hours at 4°C with dilutions of ^{125}I -labelled TNF4s (specific activity 5.7×10^6 cpm per picomole) in the presence or absence of 20-fold excess unlabelled zTNF4s, washed and measured in a gamma counter.

Received 7 February; accepted 22 March 2000.

- Moore, P. A. *et al.* BLYS: Member of the tumor necrosis factor family and B lymphocyte stimulator. *Science* **285**, 260–263 (1999).
- Scheider, P. *et al.* BAFF, a novel ligand of the tumor necrosis factor family, stimulates B cell growth. *J. Exp. Med.* **189**, 1747–1756 (1999).
- Shu, H., Hu, W. & Johnson, H. TALL-1 is a novel member of the TNF family that is down-regulated by mitogens. *J. Leukoc. Biol.* **65**, 680–683 (1999).
- Mukhopadhyay, A., Ni, J., Zhai, Y., Yu, G. & Aggarwal, B. Identification and characterization of a novel cytokine, THANK, a TNF homologue that activates apoptosis, nuclear factor- κB , and c-Jun NH_2 -terminal kinase. *J. Biol. Chem.* **274**, 15978–15981 (1999).
- Mackay, F. *et al.* Mice transgenic for BAFF develop lymphocytic disorders along with autoimmune manifestations. *J. Exp. Med.* **190**, 1697–1710 (1999).
- von Bülow, G.-U. & Bram, R. J. NF-AT activation induced by a CAML-interacting member of the tumor necrosis factor receptor superfamily. *Science* **278**, 138–141 (1997).
- Laabi, Y. *et al.* A new gene, BCM, on chromosome 16 is fused to the interleukin 2 gene by a t(4;16)(q26;p130) translocation in a malignant T cell lymphoma. *EMBO J.* **11**, 3897–3904 (1992).
- Maldry, C. *et al.* The characterization of murine BCMA gene defines it as a new member of the tumor necrosis factor receptor superfamily. *Int. Immunol.* **10**, 1693–1702 (1998).
- Smith, C. A., Farrah, T. & Goodwin, R. G. The TNF-receptor superfamily of cellular and viral proteins: activation, costimulation, and death. *Cell* **76**, 959–962.
- Bodrug, S. E. *et al.* Cyclin D1 transgene impedes lymphocyte maturation and collaborates in lymphomagenesis with the *myc* gene. *EMBO J.* **13**, 2124–2130 (1994).
- Hu, J. *et al.* An evaluation of the potential to use tumor-associated antigens as targets for antitumor T cell therapy using transgenic mice expressing a retroviral tumor antigen in normal lymphoid tissues. *J. Exp. Med.* **177**, 1681–1690 (1993).
- Hardy, R. R., Carmack, C. E., Li, Y. S. & Hayakawa, K. Distinctive developmental origins and specificities of murine CD5⁺ B cells. *Immunol. Rev.* **137**, 91–95 (1994).
- Foster, M. H. Relevance of systemic lupus erythematosus nephritis animal models to human disease. *Semin. Nephrol.* **19**, 12–24 (1999).
- Cohen, P. L. & Eisenberg, R. A. *Lpr* and *gld*: single gene models of systemic autoimmunity and lymphoproliferative disease. *Annu. Rev. Immunol.* **9**, 243–269 (1991).
- Gras, M.-P. *et al.* BCMAp: an integral membrane protein in the Golgi apparatus of human mature B lymphocytes. *Int. Immunol.* **7**, 1093–1106 (1995).
- Kotzin, B. L. Systemic lupus erythematosus. *Cell* **85**, 303–306 (1996).
- Chan, O. *et al.* A novel mouse with B cells but lacking serum antibody reveals an antibody-independent role for B cells in murine lupus. *J. Exp. Med.* **189**, 1639–1647 (1999).
- Raymond, C. K. *et al.* Development of the methylotrophic yeast *Pichia methanolica* for the expression of the 65 kilodalton isoform of human glutamate decarboxylase. *Yeast* **14**, 11–23 (1998).
- Seeburg, P. H. The human growth hormone gene family: nucleotide sequences show recent divergence and predict a new polypeptide hormone DNA. *DNA* **1**, 239–249 (1982).
- Wofsy, D. & Seaman, W. E. Successful treatment of autoimmunity in NZB/NZW F1 mice with monoclonal antibody to L3T4. *J. Exp. Med.* **161**, 378–391 (1985).
- Davis, S. *et al.* Isolation of angiopoietin-1, a ligand for the TIE2 receptor, by secretion-trap expression cloning. *Cell* **87**, 1161–1169 (1996).

Acknowledgements

We thank D. Wofsy, R. Bram and A. Nelson for helpful discussions; K. Waggie, L. Wilcox, B. Hansen, J. Lenox, C. Bosnick, S. Bayna and M. Caputo for generation and analysis of transgenic animals; A. Thorstrud, P. Shea, T. Bukowski, N. Hamacher, M. Stamm, K. De Jongh, K. Swiderek and J. Forstrom for protein purification and analysis; J. Volpone and S. McMillen for generation of antibody reagents and assay development; and C. Brandt for binding studies using zTNF4.

Correspondence and requests for materials should be addressed to J.G. (e-mail: grossj@zgi.com).

Transmembrane phosphoprotein Cbp regulates the activities of Src-family tyrosine kinases

Masahiro Kawabuchi⁺, Yoshinori Satomi[†], Toshifumi Takao[‡], Yasutsugu Shimonishi[†], Shigeyuki Nada[‡], Katsuya Nagai^{*}, Alexander Tarakhovsky[§] & Masato Okada^{*||}

Divisions of ⁺Protein Metabolism and [†]Organic Chemistry, Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan

[‡]Department of Cell Membrane Biology, Division of Biological Science, Institute of Scientific and Industrial Research, Osaka University, 8-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

[§]Laboratory for Lymphocyte Signalling, Institute for Genetics, University of Cologne, Weyertal 121, D-50931 Cologne, Germany

^{||}These authors contributed equally to this work.

The Src family of protein tyrosine kinases (Src-PTKs) is important in the regulation of growth and differentiation of eukaryotic cells. The activity of Src-PTKs in cells of different types is negatively controlled by Csk, which specifically phosphorylates a conserved regulatory tyrosine residue at the carboxy-terminal tail of the Src-PTKs^{1–3}. Csk is mainly cytoplasmic and Src-PTKs are predominantly membrane-associated. This raises a question about the mechanism of interaction between these enzymes. Here we present Cbp—a transmembrane phosphoprotein that is ubiquitously expressed and binds specifically to the SH2 domain of Csk. Cbp is involved in the membrane localization of Csk and in the Csk-mediated inhibition of c-Src. In the plasma membrane Cbp is exclusively localized in the GM1 ganglioside-enriched detergent-insoluble membrane domain, which is important in receptor-mediated signalling^{4–8}. These findings reveal Cbp as a new component of the regulatory mechanism controlling the activity of membrane-associated Src-PTKs.

The ability of Csk to relocate from the cytosol to the plasma membrane⁹ and the involvement of the SH2 and/or SH3 domains of Csk in inhibiting of Src-PTKs^{10,11} indicate the existence of a plasma-membrane-associated, Csk-binding molecule that mediates the

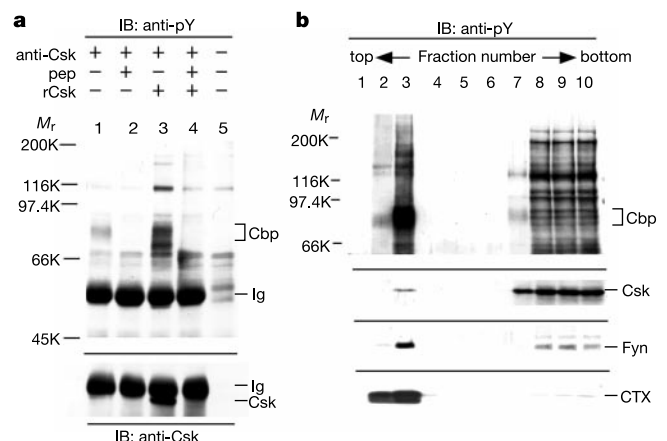


Figure 1 Csk association with the DIM-localized phosphoprotein. **a**, Association of Csk with M_r 80K–90K phosphoprotein in rat brain lysate. The lysate was incubated with anti-Csk, with or without recombinant Csk (rCsk) and Csk peptide (pep.). The immunoprecipitates (IP) were analysed by immunoblotting (IB) with anti-pY or anti-Csk (bottom). **b**, Predominant localization of Cbp in the DIM fraction. The Triton X-100 containing brain lysates were fractionated¹². Aliquots of the fractions were resolved by SDS-PAGE and analysed by immunoblotting with anti-pY, anti-Csk, anti-Fyn and the B-subunit of cholera toxin (CTX).

recruitment of Csk to Src-PTK. We used neonatal rat brain enriched in Csk and Src-PTKs as a potential source of such a molecule. Immunoprecipitation of the Csk from the brain lysate resulted in co-precipitation of a tyrosine-phosphorylated protein (Csk binding protein, Cbp) migrating as a diffuse band with a relative molecular mass of around 80,000–90,000 ($M_r \approx 80\text{K}–90\text{K}$) (Fig. 1a, lane 1). Addition of the recombinant Csk protein (rCsk) enhanced the yield of the Cbp (Fig. 1a, lane 3) and other known Csk-interacting proteins such as Fak and Paxillin¹⁰ (data not shown). However, only Cbp remained tightly bound to Csk in the presence of increased NaCl (1 M) concentrations (data not shown). While optimizing the purification conditions for Cbp, we found this protein to be

enriched in the Triton X-100 insoluble fraction of brain lysate and the degree of Cbp phosphorylation to be enhanced by the addition of MgCl_2 to the lysis buffer. Subfractionation of Triton X-100 cell lysates on a discontinuous density sucrose gradient revealed a predominant localization of phosphorylated Cbp within the detergent-insoluble membrane (DIM) fraction^{4,5} enriched for the cholera toxin (CTX) B-subunit binding ganglioside GM1 and Src-PTK Fyn (Fig. 1b, fractions 2, 3). The localization of Cbp in the DIM fraction and its interaction with the SH2 domain of Csk (data not shown) indicated a scheme for the purification of Cbp. This included isolating the DIM fraction from neonatal rat brain, dissolving it in octyl-D-glucoside, and purifying it by affinity chromatography on a resin coupled to the glutathione *S*-transferase (GST)–Csk SH2 fusion protein (Fig. 2a). The Cbp was purified further by SDS–polyacrylamide gel electrophoresis (SDS–PAGE), excised from the gel and subjected to an in-gel digestion for microsequencing.

On the basis of the amino-acid sequence of the Cbp-derived peptides, we used degenerate oligonucleotides to generate a putative complementary DNA for Cbp and then deduced the primary structure of an open reading frame (ORF) (Fig. 2b). The ORF

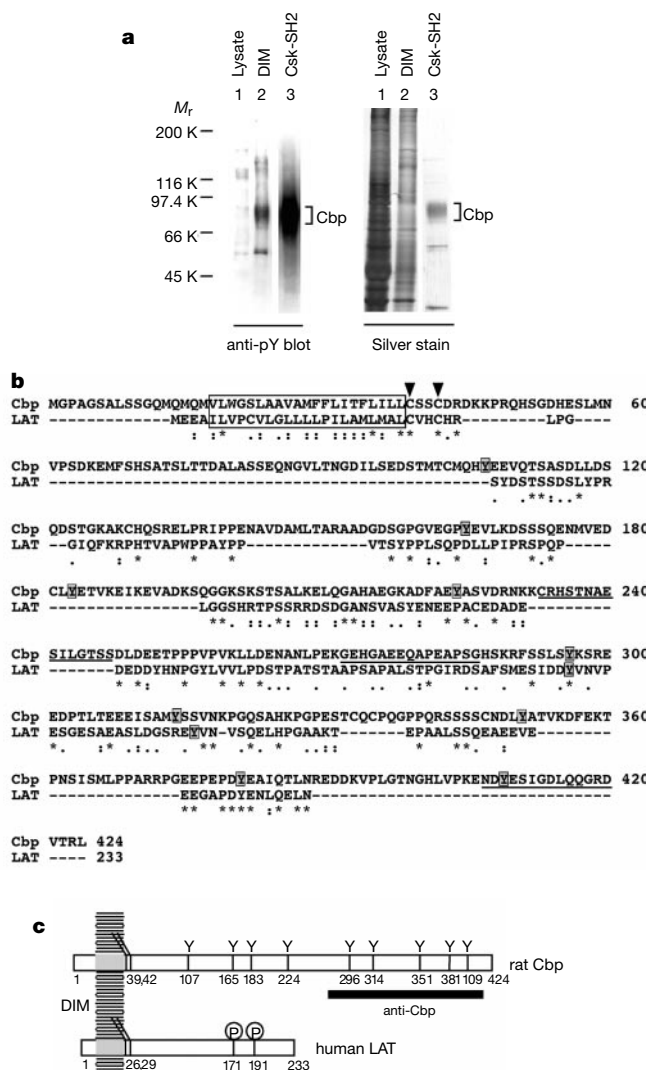


Figure 2 Purification and structure of Cbp. **a**, Cbp was purified by affinity chromatography of the DIM fraction on the resin coupled to GST–Csk SH2 (Csk-SH2). Samples from different purification steps were analysed by immunoblotting with anti-pY (left) and protein silver staining (right). **b**, Amino-acid sequences of rat Cbp and human LAT¹¹. Putative transmembrane domain is boxed. Arrowheads point to the putative palmitoylated cysteine residues¹¹. Tyrosine residues of Cbp and essential tyrosine residues of LAT¹¹ are framed. Amino-acid sequences of Cbp determined by microsequencing are underlined. Amino-acid identity, strong similarity and weaker similarity between rat Cbp and human LAT are indicated by asterisks, colons and dots below the LAT sequence, respectively. **c**, Domain structures of rat Cbp and human LAT protein. Amino-acid residues 39 and 42, and 26 and 29, are the putative palmitoylation sites of Cbp and LAT, respectively. Y, tyrosine; P, tyrosine phosphorylation sites; DIM, the detergent-insoluble membrane microdomain. A region used for the generation of anti-Cbp antibody is underlined.

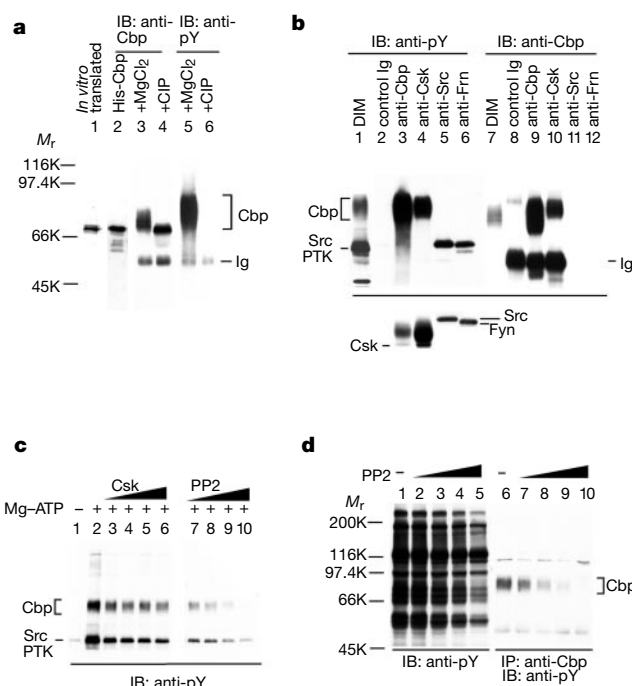


Figure 3 Characterization of Cbp protein. **a**, *In vitro* translated Cbp was resolved by SDS–PAGE and visualized by autoradiography (lane 1). Bacterially expressed His-tagged Cbp was detected by immunoblotting with anti-Cbp antibody (lane 2). Cbp was immunoprecipitated from DIM fraction prepared in the presence of MgCl_2 (lane 3), and treated with alkaline phosphatase (CIP) (lane 4). The Cbp was detected as above and the phosphorylation of immunoprecipitated Cbp was analysed by immunoblotting with anti-pY antibody (lanes 5, 6). **b**, Interaction of Csk with phosphorylated Cbp in DIM fraction. The solubilized DIM fraction was immunoprecipitated with nonimmune IgG, anti-Cbp, anti-Csk, anti-Src or anti-Fyn. The DIM fraction and the immunoprecipitates were analysed by immunoblotting with anti-pY, anti-Cbp, anti-Csk, anti-Src or anti-Fyn (lower right). **c**, Phosphorylation of Cbp *in vitro*. The solubilized DIM fraction of rat brain prepared in the absence of MgCl_2 and Na_2VO_4 was incubated with 50 μM ATP, 10 mM MgCl_2 and increasing amounts of rCsk (0, 10, 20, 40 and 80 ng, lanes 2–6) or PP2 (1, 2, 5 and 10 μM , lanes 7–10). The phosphorylation of Cbp was analysed by immunoblotting with anti-pY. **d**, Suppression of Cbp phosphorylation by a Src-PTK inhibitor *in vivo*. NIH3T3 cells were cultured in the presence of increasing amounts of PP2 (0, 1, 2, 5 and 10 μM). After incubation for 10 h, cell lysate (left panel) and the immunoprecipitated Cbp (right panel) were subjected to immunoblotting with anti-pY. Cbp protein in the immunoprecipitate was confirmed by immunoblotting with anti-Cbp (data not shown).

contains 424 amino acids with a calculated M_r of 45,912. The deduced sequence shows weak but significant homology to LAT, a linker protein expressed in T lineage cells and exclusively localized within the DIM compartment^{12,13}. The domain structure of Cbp is highly reminiscent of LAT (Fig. 2c); both of the proteins possess a short extracellular domain (Cbp, residues 1–17; LAT, residues 1–4) followed by the putative transmembrane domain (Cbp, residues

18–38; LAT, residues 5–27), putative palmitoylation sites (Cbp, residues 39 and 42; LAT, residues 26–29), and a long cytoplasmic domain containing potential sites for tyrosine phosphorylation (Fig. 2b, c). Immunoblot analysis of protein extracts derived from various rat tissues and lymphoid cell populations with anti-Cbp antibody revealed ubiquitous expression of Cbp with the highest expression in the developing brain, lung, thymus, spleen and testis (data not shown).

Translation of the Cbp messenger RNA *in vitro* or its expression in bacteria generates unphosphorylated protein with $M_r \approx 68K$ (Fig. 3a, lanes 1, 2). The difference between the deduced molecular size of Cbp (M_r 46K) and the mobility of unphosphorylated Cbp on the SDS–PAGE could be explained by the abundance of negatively charged amino acids (Asp 30 and Glu 41). The relative molecular mass of the phosphorylated Cbp precipitated from DIM fraction with an anti-Cbp antibody varies between 70K and 90K (Fig. 3a, lanes 3, 5), which could be explained by the presence of the multiple phosphorylation sites within the cytoplasmic portion of Cbp. Indeed, the treatment of the Cbp immunoprecipitate with alkaline phosphatase yields a single Cbp band of M_r 68K (Fig. 3a, lanes 4, 6). The interaction between Csk and Cbp in the DIM fraction was confirmed by immunoprecipitation assay using anti-Csk and anti-Cbp antibodies (Fig. 3b, lanes 3, 4, 9, 10). In contrast, neither Src nor Fyn associate with Cbp (Fig. 3b, lanes 5, 6, 11, 12). These data support the high selectivity of the Csk–Cbp interaction. The *in vitro* kinase assay with a DIM fraction showed strong Mg- and ATP-dependent phosphorylation of Cbp, suggesting the co-localization of Cbp and its kinase within the DIM fraction (Fig. 3c, lanes 1, 2). The nature of Cbp phosphorylating kinases remains elusive. The phosphorylation of Cbp is suppressed, albeit not completely, by rCsk *in vitro*, and by PP2, a selective inhibitor for Src-PTKs¹⁴, both *in vitro* and *in vivo* (Fig. 3c, d). This indicates the involvement of the Src-PTKs in the phosphorylation of Cbp. However, the lack of association of Cbp with Src-PTKs (Fig. 3b) and the significant, but partial, suppression of Cbp phosphorylation by Csk (Fig. 3c) indicate that kinases other than Src-PTKs may be involved in Cbp phosphorylation.

Co-expression of exogenous c-Src and wild-type Cbp in 293T cells, characterized by barely detectable endogenous c-Src activity (data not shown), resulted in Cbp phosphorylation and its association with Csk, as revealed by the presence of phosphorylated Cbp in immunoprecipitates with anti-Csk antibodies (Fig. 4A, lane 1–3).

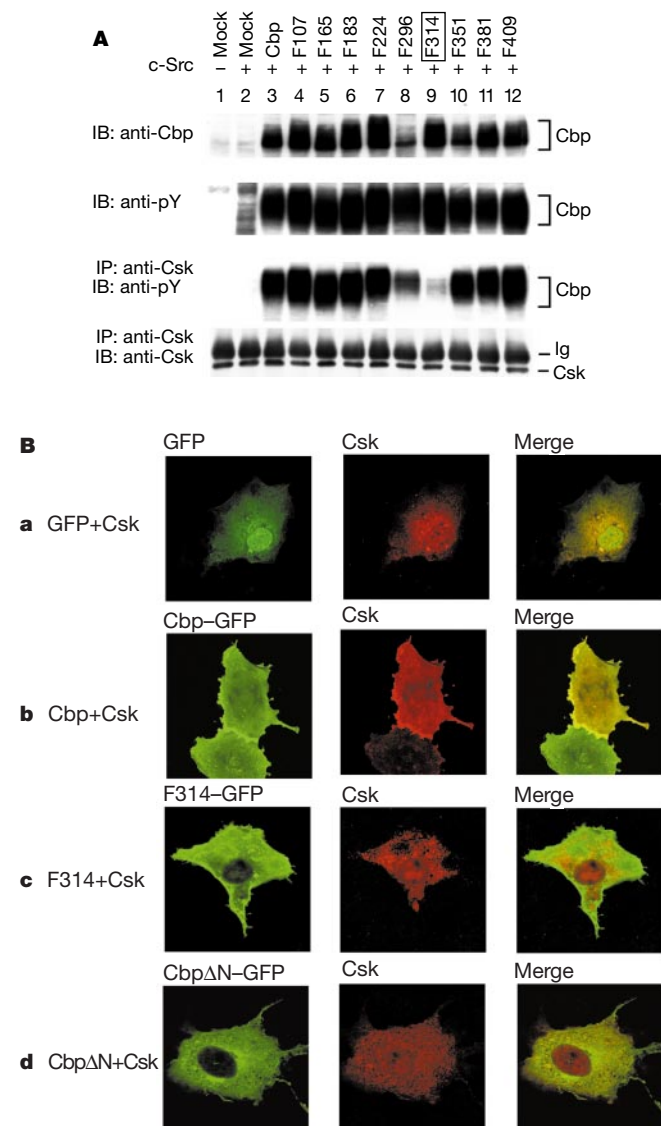


Figure 4 Association and co-localization of Cbp with Csk in COS7 cells. **A**, Identification of Csk binding site in Cbp. The exogenous wild-type and mutated Cbp were co-expressed with c-Src in 293T cells. The expression of Cbp protein and its tyrosine phosphorylation were detected by immunoblotting with anti-Cbp and anti-pY, respectively. Tyrosine phosphorylated Cbp bound to Csk was analysed by immunoblotting of the Csk immunoprecipitates (IP) with anti-pY. The amount of Csk within the Csk immunoprecipitates was controlled by immunoblotting with anti-Csk. **B**, Co-localization of Cbp and Csk in COS7 cells. COS7 cells were transiently co-transfected with the pME18Sneo–Csk in combination with the pEGFP–Csk (a), pEGFP–Cbp (b), pEGFP–F314 (c) or pEGFP–CbpΔN (d) expression vectors. After 24 h, the cells were fixed and the Csk was revealed by incubation of the fixed cells with anti-Csk antibody followed by staining with Rhodamin-conjugated goat anti-rabbit antibody. Areas of Csk (red) and GFP–Cbp (green) co-localization are yellow (Merge). The expression of endogenous Cbp in COS7 cells was negligibly low in comparison with levels achieved by expression of exogenous GFP–Cbp (data not shown).

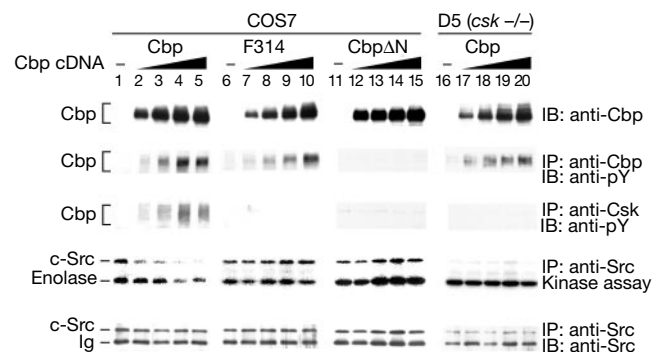


Figure 5 Overexpression of the Cbp suppresses the kinase activity of c-Src in COS7 cells. Increasing amounts of pCMV–Cbp, pCMV–F314 or pCMV–CbpΔN were transfected into the COS7 cells (5×10^5). Total transfected DNA was made up to 4 μ g with pCMV–tag. pCMV–Cbp was also transfected into D5 cells (csk^{-/-})¹⁵. Transfected cell lysates were incubated with anti-Cbp or anti-Csk. Immunoprecipitates were analysed by immunoblotting with anti-Cbp and anti-pY. Endogenous c-Src was immunoprecipitated from the lysates and the kinase activity was estimated by *in vitro* kinase assay. The c-Src protein in the immunoprecipitate was detected by immunoblotting with anti-Src. The amount of phosphorylated enolase and autophosphorylated c-Src was quantified by a Bioimage Analyser.

In 293T cells, expression of mutated forms of Cbp with individual tyrosine residues replaced by phenylalanine revealed the critical importance of Tyr 314 in Cbp binding to Csk (Fig. 4A, lane 9). Furthermore, Tyr 314 is essential for the co-localization of Cbp and Csk at the plasma membrane. We expressed exogenous Csk alone, or with the wild-type or mutated Cbp–green fluorescent fusion protein (Cbp–GFP) in COS7 cells. In the absence of Cbp, the exogenous Csk was diffused throughout the cytoplasm of transfected cells (Fig. 4Ba). When exogenous Csk was co-expressed with Cbp–GFP, most Csk was co-localized with Cbp–GFP at the cytoplasmic face of the plasma membrane (Fig. 4Bb). The Cbp with Phe instead of Tyr 314 (F314) or Cbp lacking its amino-terminal membrane-associated region (CbpΔN) did not co-localize with Csk at the plasma membrane (Fig. 4Bc, d).

We analysed the effects of the Csk–Cbp interaction on the activity of Src-PTKs *in vivo* by expressing exogenous Cbp in COS7 cells. This resulted in high c-Src activity, as compared with 293T cells. (Fig. 5a, and data not shown). The levels of expression and phosphorylation of Cbp correlated directly with the amounts of the Cbp-bound, phosphorylated Csk, but were in inverse proportion to the activity of the endogenous c-Src. The mutated Cbp (F314) was phosphorylated but could neither bind to Csk nor affect the activity of c-Src. The CbpΔN was not tyrosine-phosphorylated or associated with Csk, and did not affect the activity of c-Src. Finally, expression of Cbp in a Csk-deficient fibroblast cell line (D5) (ref. 15) did not affect the expression and activity of c-Src in these cells.

Our data indicate the existence of a negative-feedback signalling loop in which the activation of Src-PTKs, and possibly other tyrosine kinases, results in the phosphorylation of Cbp followed by recruitment of Csk to the membrane and suppression of the Src-PTKs. This process is likely to be reversible. In this context, a tyrosine phosphatase such as SHP-2 (ref. 16) or PTPα (ref. 17) may dephosphorylate Cbp leading to displacement of Csk and reactivation of Src-PTKs. In addition, the ubiquitin-dependent degradation of activated Src-PTKs¹⁸ may lead to reduction in phosphorylation of Cbp followed by its uncoupling from Csk. The exclusive localization of Cbp in the DIM fraction indicates a unique involvement of this membrane domain in dynamic regulation of the activity of ligand-activated Src-PTKs. Finally, the ubiquitous expression of Cbp in various cell types indicates its importance in the regulation of physiological processes. □

Methods

Antibodies and cells

We generated polyclonal rabbit antibodies against a peptide corresponding to the C-terminal 14 amino acids (residues 437–450) of Csk (anti-Csk) and purified them by affinity to resin coupled to the peptide. We generated anti-Cbp against the GST–Cbp (residues 275–419) fusion protein and purified on the GST–Cbp-coupled resin. We obtained anti-Fyn, anti-Fak and anti-Paxillin antibodies from Transduction Laboratories, anti-Src antibody and PP2 from Calbiochem Novabiochem, anti-phosphotyrosine antibody (anti-pY, 4G10) from Upstate Biotechnology, and horseradish peroxidase (HRP)-conjugated B-subunit of cholera toxin (CTX) from Sigma. We did immunoblotting as described¹⁹. We cultured 293T cells, COS7 cells and NIH3T3 cells in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum.

Immunoprecipitation and subcellular fractionation

We homogenized tissues and cells in ODG buffer (50 mM Tris-HCl (pH 8), 1 mM EDTA, 0.25 M NaCl, 20 mM NaF, 1 mM Na₂VO₄, 2% Octyl-β-glucoside, 5 mM β-mercaptoethanol, 10 μg ml⁻¹ aprotinin, 10 μg ml⁻¹ leupeptin, 1 mM PMSF, 5% glycerol) at a ratio of 8:1 (volume to weight). After incubation of the homogenate with Protein G-Sepharose (Amersham Pharmacia) for 1 h, we cleared it by centrifugation at 20,000g for 20 min. We incubated the supernatant for 1 h with the relevant antibody and Protein G-Sepharose. We washed the immunoprecipitates with ODG buffer and analysed them by immunoblotting.

For the analysis of protein association with DIM fraction, we homogenized tissues in Triton X-100 lysis buffer (50 mM Tris-HCl (pH 8.0), 10 mM MgCl₂, 0.15 M NaCl, 20 mM NaF, 1 mM Na₂VO₄, 1% Triton X-100, 5 mM β-mercaptoethanol, 10 μg ml⁻¹ aprotinin, 10 μg ml⁻¹ leupeptin, 1 mM PMSF, 5% glycerol) at a ratio of 8:1 (volume to weight). After stirring for 1 h at 4 °C, we fractionated the homogenate on a small-scale sucrose gradient similar to one described below (4 ml, 40% sucrose:5 ml, 35% sucrose:1 ml, 5% sucrose) We centrifuged this in a Beckman SW41 rotor for 12 h at 200,000g, 4 °C. We collected 1.0-ml fractions from the top of the gradient and analysed them by immunoblotting.

Purification of Cbp, peptide analysis and cDNA cloning

We homogenized brain tissue derived from 1 day-old 250 g rats in Triton X-100 lysis buffer at a ratio of 8:1 (volume to weight). We stirred the homogenate for 1 h at 4 °C and then mixed 15 ml per tube with an equal volume of 80% sucrose in Buffer A (50 mM Tris-HCl (pH 7.4), 50 mM NaCl, 10 mM MgCl₂, 1 mM Na₂VO₄). After transfer of the mixture to the centrifuge tube, we sequentially overlaid it with 25 ml of 35% sucrose in Buffer A and 5 ml of 5% sucrose in Buffer A. After centrifugation for 3 h at 100,000g, 4 °C, in a Beckman 550 rotor, we collected the material localized at the interface between 35% and 5% sucrose. This we diluted threefold with Buffer A, and centrifuged for 1 h at 100,000g, 4 °C. We used ODG buffer to dissolve the pellet, and then passed the cleared supernatant through an affinity column coupled to GST–Csk SH2. We eluted the bound material with 1% SDS in 25 mM Tris-HCl (pH 6.8). After concentrating the sample under vacuum, we separated it by SDS–PAGE. We excised the M_r 80K–90K Csk-binding protein from the SDS–PAGE gel and digested it in the gel slice with lysyl-endopeptidase²⁰. We used one-tenth of the recovered peptides for peptide-mass fingerprinting with matrix-assisted laser desorption ionization mass spectrometry (MALDI-TOFMS), and the rest for separation by HPLC. We sequenced three peptides by gas phase microsequencing, and used the sequences to design degenerate oligonucleotide primers for PCR. We used random hexamers with poly(A)⁺ RNA from neonatal rat brain to synthesize the template cDNA. A PCR between the primers 25.6-2-F, 5'-GGGCGGARGARGARGARGCCNGA-3', and 62.6-C-R; 5'-CGNCCYGTGTGNARRTCNCCDAT-3', yielded a specific product of 435 base pairs (bp). We isolated the full-length Cbp cDNA by the 5'- and 3'-rapid amplification of cDNA ends method using Marathon cDNA amplification Kit (Clontech). We deduced the amino-acid sequences of the putative ORF from the cDNA sequence. We evaluated the potential post-translational modifications of Cbp by comparing the peptide-mass fingerprint from the Cbp with that of predicted primary structure from the cDNA by MS-Match, a software aid for protein identification by MS (<http://www.protein.osaka-u.ac.jp/organic>).

Protein expression *in vitro* and *in vivo*

We cloned the Cbp cDNA in pBluscript II (Stratagene) and *in vitro* translated it in the presence of [³⁵S] methionine using a TNT-coupled reticulocyte system (Promega). We expressed His-tagged Cbp protein in *Escherichia coli* DH5α transformed with pTrcHis vector (Invitrogen) carrying Cbp cDNA. Transfections of 293T cells and COS7 cells used lipofectamine2000 (Gibco BRL) or FuGENE6 (Boehringer Mannheim). We expressed Cbp (CbpΔN, residues 53–424) and GFP-tagged Cbp (CbpΔN) using pCMV-tag (Stratagene) and pEGFP-N1 vectors (Clontech), respectively. We used a PCR-based procedure to generate the Cbp mutants. We expressed Csk and c-Src using pME18Sneo (a gift from K. Maruyama) and pCDNA3 vectors (Invitrogen), respectively. We purified the recombinant Csk from Sf9 cells transfected with Baculovirus vector carrying Csk cDNA (unpublished).

In vitro kinase assay

We extracted the DIM fraction in the ODG buffer and incubated the cleared lysates in a kinase buffer (50 mM Tris-HCl (pH 7.4), 10 mM MgCl₂) with 50 μM ATP. After incubation for 10 min at 30 °C, we terminated the reaction by adding the SDS sample buffer. We analysed the protein phosphorylation by immunoblotting with anti-pY antibody. For the analysis of the Src-mediated substrate phosphorylation, we incubated immunoprecipitated c-Src with enolase in the presence of 10 μM [γ-³²P]ATP, and quantified the amount of phosphorylation using BAS2000 Bioimage analyser (Fuji).

Immunofluorescence staining and confocal microscopy

We grew COS7 cells transfected with the pEGFP-N1, pEGFP-Cbp, pEGFP-F314 and pEGFP-CbpΔN on glass coverslips for two days before fixation. We grew cells transfected with pME18Sneo-Csk overnight before antibody staining. We stained Csk by immunofluorescence²¹ using Rhodamin-conjugated anti-rabbit IgG antibody (Amersham Pharmacia Biotech). We observed the specimens by confocal microscopy using an LSM510 (Carl Zeiss).

Received 12 January; accepted 3 March 2000.

- Nada, S., Okada, M., MacAuley, A., Cooper, J. A. & Nakagawa, H. Cloning of a complementary DNA for a protein-tyrosine kinase that specifically phosphorylates a negative regulatory site of p60^{src}. *Nature* **351**, 69–72 (1991).
- Nada, S. *et al.* Constitutive activation of Src family kinases in mouse embryos that lack Csk. *Cell* **73**, 1125–1135 (1993).
- Brown, M. T. & Cooper, J. A. Regulation, substrates and function of src. *Biochem. Biophys. Acta* **1287**, 121–149 (1996).
- Simons, K. & Ikonen, E. Functional rafts in cell membranes. *Nature* **387**, 569–572 (1997).
- Brown, D. A. & London, E. Functions of lipid rafts in biological membranes. *Annu. Rev. Cell. Dev. Biol.* **14**, 111–136 (1998).
- Anderson, R. G. The caveolae membrane system. *Annu. Rev. Biochem.* **67**, 199–225 (1998).
- Xavier, R., Brennan, T., Li, Q., McCormack, C. & Seed, B. Membrane compartmentation is required for efficient T cell activation. *Immunity* **8**, 723–732 (1998).
- Montixi, C. *et al.* Engagement of T cell receptor triggers its recruitment to low-density detergent-insoluble membrane domains. *EMBO J.* **17**, 5334–5348 (1998).
- Howell, B. W. & Cooper, J. A. Csk suppression of Src involves movement of Csk to sites of Src activity. *Mol. Cell. Biol.* **14**, 5402–5411 (1994).
- Sabe, H., Hata, A., Okada, M., Nakagawa, H. & Hanafusa, H. Analysis of the binding of the Src homology 2 domain of Csk to tyrosine-phosphorylated proteins in the suppression and mitotic activation of c-Src. *Proc. Natl. Acad. Sci. USA* **91**, 3984–3988 (1994).
- Cloutier, J. E., Chow, L. M. & Veillette, A. Requirement of the SH3 and SH2 domains for the inhibitory function of tyrosine protein kinase p50^{sk} in T lymphocytes. *Mol. Cell. Biol.* **15**, 5937–5944 (1995).

12. Zhang, W., Sloan-Lancaster, J., Kitchen, J., Triple, R. P. & Samelson, L. E. LAT: the ZAP-70 tyrosine kinase substrate that links T cell receptor to cellular activation. *Cell* **9**, 83–92 (1998).
13. Zhang, W., Triple, R. P. & Samelson, L. E. LAT palmitoylation: its essential role in membrane microdomain targeting and tyrosine phosphorylation during T cell activation. *Immunity* **9**, 239–246 (1998).
14. Hanke, J. H. *et al.* Discovery of a novel, potent and Src family selective protein tyrosine kinase inhibitor. *J. Biol. Chem.* **271**, 695–701 (1996).
15. Nada, S., Okada, M., Aizawa, S. & Nakagawa, H. Identification of major tyrosine-phosphorylated proteins in Csk-deficient cells. *Oncogene* **9**, 3571–3578 (1994).
16. Tsuda, M. *et al.* Integrin-mediated tyrosine phosphorylation of SHPS-1 and its association with SHP-2. Roles of Fak and Src family kinases. *J. Biol. Chem.* **273**, 13223–13229 (1998).
17. Harder, K. W., Moller, N. P., Peacock, J. W. & Jirik, F. R. Protein-tyrosine phosphatase alpha regulates Src family kinases and alters cell-substratum adhesion. *J. Biol. Chem.* **273**, 31890–31900 (1998).
18. Hakak, Y. & Martin, G. S. Ubiquitin-dependent degradation of active Src. *Curr. Biol.* **9**, 1039–1042 (1999).
19. Kobayashi, S., Okumura, N., Okada, M. & Nagai, K. Depolarization-induced tyrosine phosphorylation of p130^{cas}. *J. Biochem. (Tokyo)* **123**, 624–629 (1998).
20. Jensen, O. N., Wilm, M., Shevchenko, A. & Mann, M. in *Methods in Molecular Biology* Vol. 112 (ed. Link, A. J.) 513–530 (Humana, Totowa, 1999).
21. Harlow, E. & Lane, D. in *Antibodies, a Laboratory Manual* 359–420 (Cold Spring Harbor Laboratory, 1988).

Acknowledgements

We thank Y. Yoshimura and S. Norioka for peptide sequence analysis; S. Yoshimura and S. Aimoto for peptide synthesis; K. Nakamura and H. Sabe for technical advice; Y. Takayama and T. Hirose for production of recombinant Csk in Sf9 cells; and C. Schmedt for critical comments. This work was supported by the Ministry of Education, Science, Sports and Culture of Japan and the Program for Promotion of Fundamental Studies in Health Sciences of the Organization for Pharmaceutical Safety and Research. A.T. was supported by the Deutsche Forschungsgemeinschaft and the Fonds der Chemie.

Correspondence and requests for materials should be addressed to M.O. (e-mail: okadam@protein.osaka-u.ac.jp) or A.T. (e-mail: sasha@mac.genetik.uni-koeln.de). The Cbp sequence has been deposited in the DDBJ under accession code AB038038.

CpG methylation is maintained in human cancer cells lacking DNMT1

Ina Rhee^{†‡}, Kam-Wing Jair[†], Ray-Whay Chiu Yen[†], Christoph Lengauer[†], James G. Herman[†], Kenneth W. Kinzler^{†‡}, Bert Vogelstein^{†‡}, Stephen B. Baylin^{†‡} & Kornel E. Schuebel[†]

* The Howard Hughes Medical Institute, † The Johns Hopkins Oncology Center, and ‡ The Program in Human Genetics, Johns Hopkins University School of Medicine, 1650 Orleans Street, Baltimore, Maryland 21231, USA

Hypermethylation is associated with the silencing of tumour susceptibility genes in several forms of cancer^{1,2}; however, the mechanisms responsible for this aberrant methylation are poorly understood^{3,4}. The prototypic DNA methyltransferase, DNMT1, has been widely assumed to be responsible for most of the methylation of the human genome, including the abnormal methylation found in cancers^{5,6}. To test this hypothesis, we disrupted the *DNMT1* gene through homologous recombination in human colorectal carcinoma cells. Here we show that cells lacking *DNMT1* exhibited markedly decreased cellular DNA methyltransferase activity, but there was only a 20% decrease in overall genomic methylation. Although juxtacentromeric satellites became significantly demethylated, most of the loci that we analysed, including the tumour suppressor gene *p16^{INK4a}*, remained fully methylated and silenced. These results indicate that DNMT1 has an unsuspected degree of regional specificity in human cells and that methylating activities other than DNMT1 can maintain the methylation of most of the genome.

An efficient strategy for generating somatic cell knockouts through homologous recombination⁷ was used to disrupt both alleles of *DNMT1* in the colorectal carcinoma line, HCT116. We chose HCT116 for these experiments because of its propensity to

silence genes through promoter hypermethylation^{8,9} and its stable karyotype¹⁰. Homologous targeting replaced exons 3, 4 and 5 of *DNMT1* with a promoterless hygromycin resistance gene, which is expressed from the endogenous *DNMT1* promoter as a fusion protein containing the first 43 amino acids of DNMT1 (Fig. 1a). Genotypes were confirmed by Southern blotting, and loss of the *DNMT1* gene product was verified by western blotting using amino- and carboxy-terminal antibodies (Fig. 1b, c). We obtained three independent *DNMT1*^{+/-} and three *DNMT1*^{-/-} clones for these studies. All clones of the same genotype were found to behave identically.

The *DNMT1*^{-/-} cells exhibited normal morphology and have survived over 75 passages in culture (300 cell generations), although

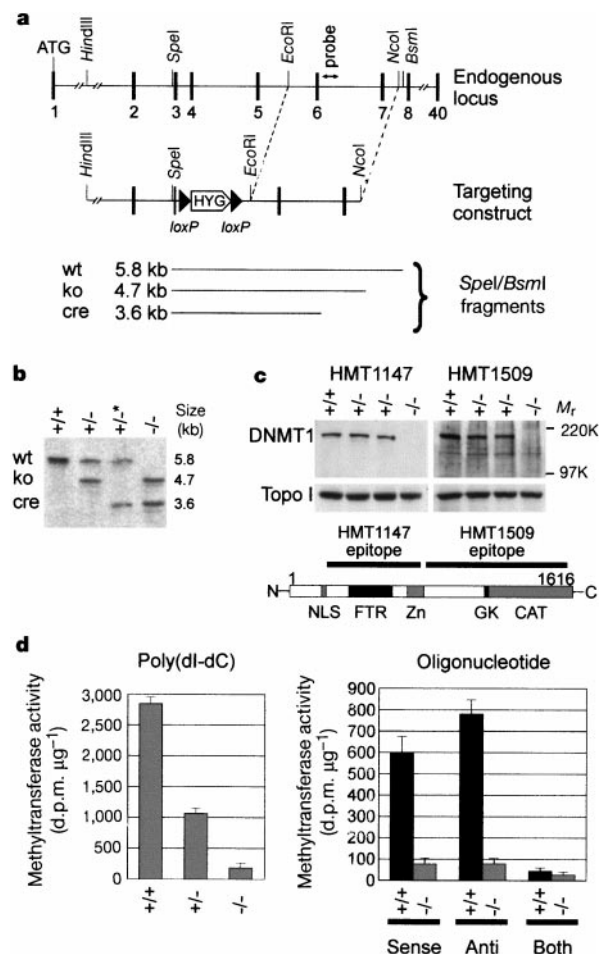


Figure 1 Generation of human cells deficient in *DNMT1*. **a**, Map of the human *DNMT1* locus and targeting construct. Numbered boxes represent exons. HYG denotes the hygromycin resistance gene. The Southern blotting probe shown detects *SpeI/BsmI* fragments from the wild-type (wt) allele, targeted allele (ko), and targeted allele after Cre–loxP excision (cre) of the indicated size. **b**, Southern blot analysis of *SpeI/BsmI* digests of genomic DNA from cells of the indicated *DNMT1* genotypes. The *DNMT1*^{+/-} clone was derived from a *DNMT1*^{+/-} clone from which the HYG gene had been removed through cre excision. **c**, Western blot analysis using the anti-DNMT1 polyclonal antibodies HMT1147 and HMT1509 (ref. 26) or an anti-topoisomerase I control antibody. The epitopes that are recognized by HMT1147 (residues 250–687) and HMT1509 (724–1,570) are shown in relation to the nuclear localization signal (NLS), replication foci targeting sequence (FTR), zinc-binding domain (Zn), glycine–lysine repeat domain (GK) and catalytic domain (CAT) of DNMT1. **d**, DNA methyltransferase activity using either the synthetic template poly(dI-dC)-poly(dI-dC) (left panel) or a double-stranded oligonucleotide template derived from the E-cadherin CpG island (right panel)¹¹. The oligonucleotide template was methylated on either the sense or the antisense (Anti) strand, or on both strands.

they grew at a slightly slower rate than parental or *DNMT1*^{+/+} cells. To measure DNA methyltransferase activity, cell lysates were incubated with poly(dI-dC)-poly(dI-dC), and the level of incorporation of tritium from [methyl-³H]S-adenosyl methionine into this synthetic template was measured. *DNMT1*^{+/+} and *DNMT1*^{-/-} cells displayed 66 ± 4% and 96 ± 2% decreases in methyltransferase activity, respectively. Similar results were obtained when the assays were repeated using a hemi-methylated oligonucleotide template¹¹ (Fig. 1d).

Genomic 5-methylcytosine (m⁵C) was quantified in the mutant cells by reversed-phase high performance liquid chromatography (HPLC)^{12,13} after complete hydrolysis of genomic DNA to nucleosides (Fig. 2a). In contrast to the marked decrease in methyltransferase activity measured in the *DNMT1*^{-/-} cells, only a 20 ± 4% decrease in total m⁵C content occurred (Fig. 2b). This discrepancy between the level of cellular methyltransferase activity and total m⁵C was observed in each of three mutant clones in three independent assays.

The high degree of retention of m⁵C prompted us to survey the methylation status of specific genomic sequences. We carried out Southern blot analyses after digesting genomic DNA with restriction endonucleases that cleave only sequences lacking methylation at CpG residues within their recognition sites. Probes for alphoid satellite sequences at the centromeres of chromosomes 3, 4, 7, 10, 12, 17, 18, 22 and X failed to reveal any demethylation in the *DNMT1*^{-/-} cells as compared with in parental cells (Fig. 3a; and data not shown). When classical satellite 2 and 3 sequences at the juxtacentromeric regions of chromosomes 1, 9, 15 and 16 were queried, however, marked demethylation was evident in the mutant clones (Fig. 3b; and data not shown). The specificity of demethylation in *DNMT1*^{-/-} cells was confirmed through comparison with cells treated with 5-aza-2'-deoxycytidine (5-aza-dC) to globally demethylate the genome. The digestion patterns of alphoid satellite sequences were clearly different in *DNMT1*^{-/-} cells as compared with 5-aza-dC-treated cells (Fig. 3a), whereas the digestion patterns of classical satellite sequences were similar in both *DNMT1*^{-/-} cells and 5-aza-dC-treated cells (Fig. 3b). In contrast to DNA from 5-aza-dC-treated cells (Fig. 3c, d), DNA from *DNMT1*^{-/-} cells remained mostly undigested after cleavage with either *Hpa*II (Fig. 3c) or *Bst*BI (Fig. 3d), reflecting retention of extensive global methylation. Thus, the marked changes in methylation-sensitive restriction patterns seen in the juxtacentromeric sequences reflected specific, regional demethylation, rather than diffuse, global loss.

Southern blot analyses with probes for two families of widely dispersed *Alu* short interspersed elements also showed a preservation of methylation in *DNMT1*^{-/-} cells (Fig. 3e). To examine the methylation of these sequences at higher resolution, we treated DNA from *DNMT1*^{+/+} and *DNMT1*^{-/-} cells with bisulphite. Bisulphite causes deamination of unmethylated cytosines to uridine,

thereby allowing discrimination between unmethylated and methylated cytosine residues through sequence analysis¹⁴. Representative *Alu*-Sx and *Alu*-J elements were found to be equivalently methylated in wild-type and mutant cells using this technique (Fig. 3f).

To investigate the methylation of a single-copy gene, we studied the *p16*^{INK4a} gene, the prototypic tumour suppressor gene that is silenced in association with CpG-island methylation. In HCT116 cells, one allele of *p16*^{INK4a} is wild-type and the other has a truncating frameshift mutation¹⁵. The promoter of the wild-type allele is exclusively methylated in these cells, and only the mutant allele is expressed⁹. We assessed the methylation of *p16*^{INK4a} in *DNMT1*-deficient cells by sequencing bisulphite-treated DNA. Each of the 35 CpG sites within the *p16*^{INK4a} CpG island remained methylated in *DNMT1*^{-/-} cells at the ~50% frequency observed in parental cells, which is consistent with monoallelic methylation (Fig. 4a). We evaluated expression of the two *p16*^{INK4a} alleles by using polymerase chain reaction with reverse transcriptase (RT-PCR)-based methods. Sequencing of the resultant PCR products showed that only the mutant allele was expressed in the *DNMT1*^{-/-} cells, just as in the parental *DNMT1*^{+/+} cells (Fig. 4c). We also amplified the *p16*^{INK4a} gene from genomic DNA and sequenced it to confirm that the *DNMT1*^{-/-} cells retained both wild-type and mutant alleles

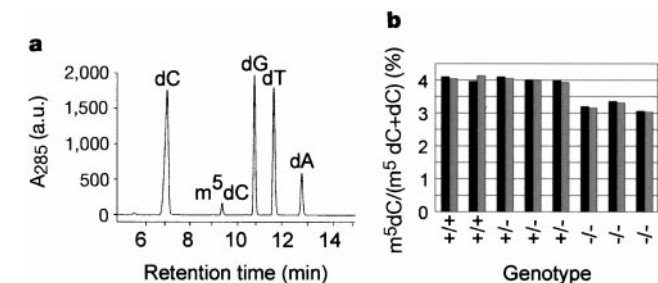


Figure 2 5-Methylcytosine (m⁵C) content of *DNMT1*^{-/-} cells. **a**, Representative chromatogram obtained upon reversed-phase HPLC separation of genomic DNA digests from wild-type HCT116 cells. **b**, Measurement of m⁵C content as a percentage of the total cytosine pool. Duplicate samples from independent clones of the indicated genotypes are shown.

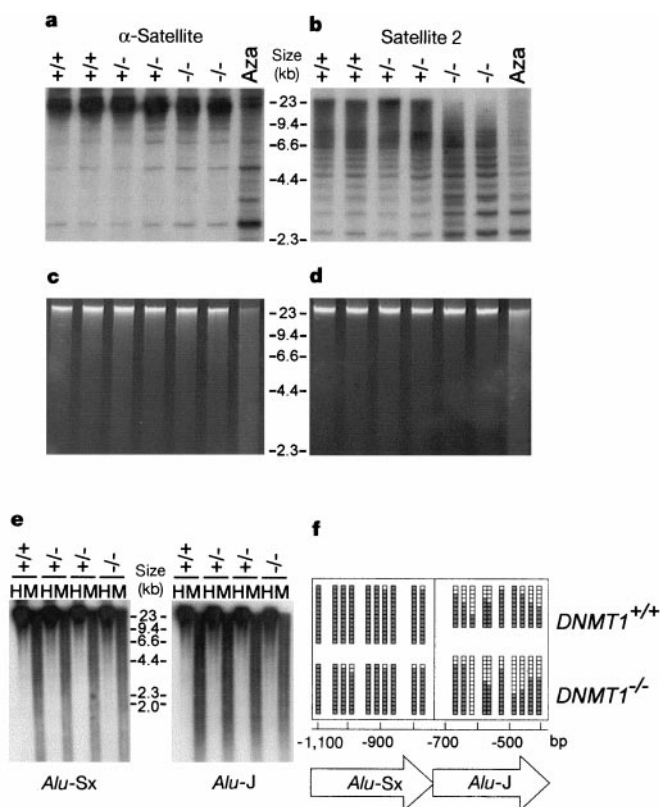


Figure 3 CpG methylation of repetitive regions. **a, b**, Southern blotting was carried out with DNA from independent clones of the indicated genotype or from wild-type cells treated with 5-aza-dC (Aza). Probes were derived from a centromeric alphoid sequence from chromosome 18 (ref. 10) (**a**) or a juxtacentromeric satellite 2 sequence from chromosome 1 (ref. 27) (**b**). Genomic DNA was digested with *Hpa*II (**a**) or *Bst*BI (**b**); these do not cleave methylated sites. **c, d**, Ethidium-bromide-stained agarose gels corresponding to the blots shown in **a** and **b**, respectively. **e**, Southern blots using probes for two different *Alu* family sequences²⁸. DNA was digested with *Hpa*II (H) or *Msp*I (M), which is an isoschizomer of *Hpa*II that is insensitive to methylation. **f**, Genomic bisulphite sequencing. The *Alu*-Sx and *Alu*-J elements within the CpG island of *E-cadherin* were amplified by PCR and cloned from bisulphite-treated DNA derived from cells of the indicated genotype. The methylation status of each CpG dinucleotide in each clone sequenced is represented as a box that is shaded if the position is methylated and white if it is not. Numbers indicate position relative to the transcriptional start site.

(Fig. 4c). Similar transcriptional and genomic analyses of *TIMP-3*, a gene whose CpG island is fully methylated on both alleles and whose expression is silenced in parental HCT116 cells¹⁶, also revealed maintenance of both methylation and silencing (Fig. 4b). We studied several additional CpG islands that are methylated in HCT116 cells, all of which retained their characteristic methylation and expression patterns in the *DNMT1*^{-/-} cells (data not shown).

To determine whether DNMT1-mediated methylation was responsible for the initiation or maintenance of exogenously introduced DNA sequences, we used a retroviral assay⁸. We infected *DNMT1*^{+/+} and *DNMT1*^{-/-} cells with a retrovirus encoding a β -galactosidase reporter gene whose expression is driven by the retroviral long terminal repeat (LTR) promoter. Previous studies have shown that the integrated retroviral β -galactosidase gene in wild-type HCT116 clones is silenced as a result of DNA methylation of the LTR⁸. Infected clones were isolated and stained with X-gal to assess β -galactosidase expression. *DNMT1*^{-/-} cells were found to retain the silencing phenotype of the parental cells, as shown by the preponderance of white colonies (Fig. 5a, b).

The fact that abolition of DNMT1 activity affects specific regions of the genome while other regions remain methylated indicated that additional methylating activities persist in the homozygous mutant

cells. We used 5-aza-dC as an independent probe for such persistent methylating activity. 5-Aza-dC is incorporated into genomic DNA where it forms a covalent complex with methyltransferase active sites. This suicide inhibition depletes methyltransferase activity, resulting in generalized demethylation^{17,18}. In addition, the resultant DNA-protein complexes constitute toxic DNA damage¹⁹. Upon treatment of *DNMT1*^{+/+} and *DNMT1*^{-/-} cells with 5-aza-dC, demethylation and corresponding re-expression of genes such as *p16*^{INK4a} and *TIMP-3* was indeed observed (Fig. 4a, b). Furthermore, the sensitivity of wild-type and homozygous mutant cells to 5-aza-dC-mediated toxicity, as measured by colony formation assays, was similar (Fig. 5c). Thus, the main target of 5-aza-dC remains intact after elimination of DNMT1.

Embryonic stem cells and embryos nullizygous for the murine *Dnmt1* gene have been studied extensively⁵ and compared with analogous murine mutants lacking one or both of the DNA methyltransferases, Dnmt3a and Dnmt3b (refs 20, 21). Only *Dnmt1* mutants show marked loss of genomic m⁵C and demethylation of a variety of surveyed sequences^{5,22,23}. Moreover, *Dnmt1* mutants have substantial resistance to 5-aza-dC toxicity¹⁹, which is consistent with the marked reduction of methyltransferase activity in these cells. Such studies have led to the idea that

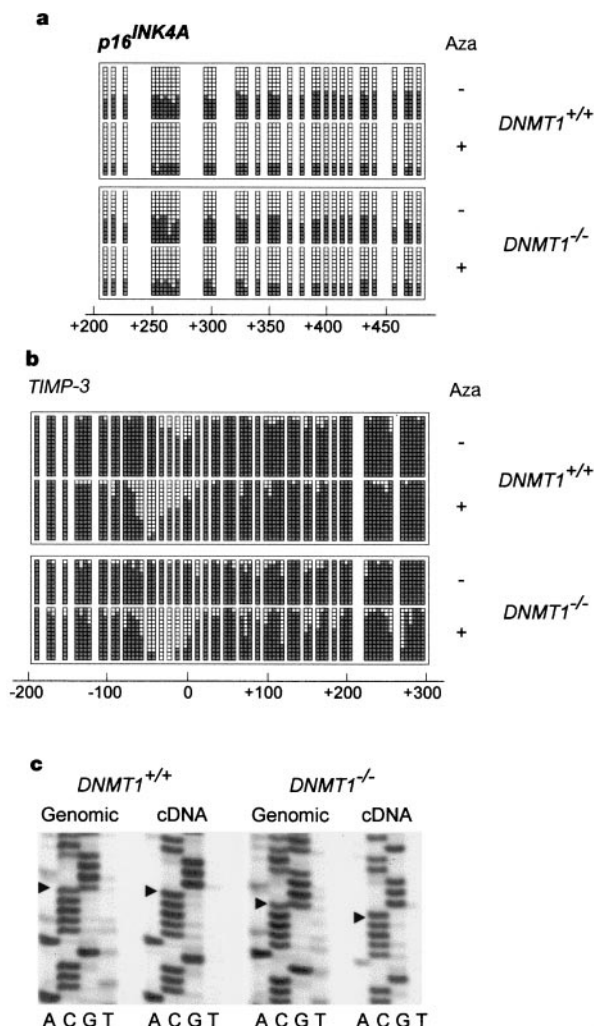


Figure 4 *DNMT1*^{-/-} cells retain the silencing phenotype of parental cells. **a, b**, Genomic bisulphite sequencing. The CpG islands of *p16*^{INK4a} (ref. 9) (**a**) and *TIMP-3* (ref. 16) (**b**) were analysed and depicted as in Fig. 3e before or after treatment with 5-aza-2'-deoxycytidine (Aza). **c**, *p16*^{INK4a} was amplified from genomic DNA or cDNA as indicated. Arrows point to the position of an insertional frameshift mutation in exon 1. Only the frameshifted allele is visible in the cDNA sequence.

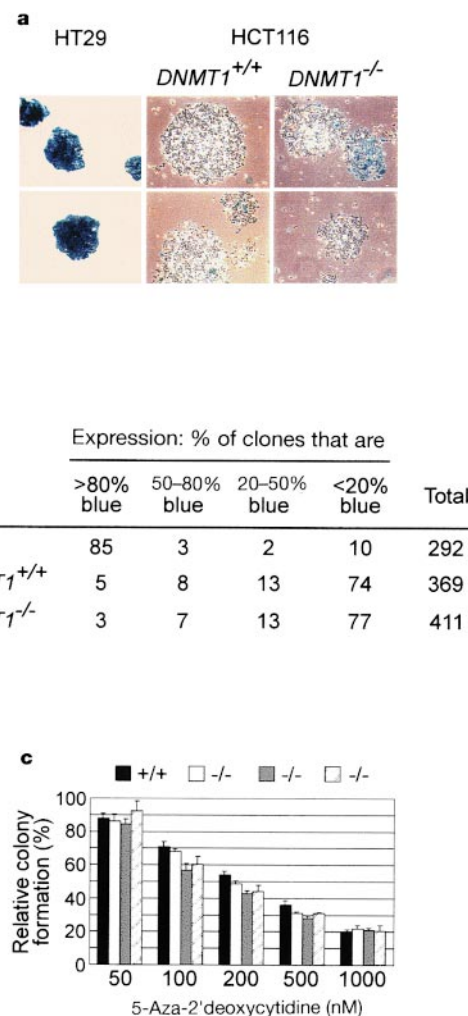


Figure 5 Silencing of exogenous sequences is maintained in *DNMT1*^{-/-} cells. **a, b**, *DNMT1*^{+/+} and *DNMT1*^{-/-} cells were infected with a retrovirus encoding β -galactosidase⁸, and resultant clones were stained with X-gal. HT29 cells served as a positive control for β -galactosidase expression. **c**, Clones were treated transiently with 5-aza-dC and allowed to form colonies in drug-free medium.

DNMT1 is responsible for maintaining most of the methylation in mammalian cells; however, our biochemical, genetic and pharmacological data indicate that methyltransferases other than DNMT1 are sufficient to maintain most DNA methylation in HCT116 human cells. Of the other known mammalian methyltransferases, *DNMT3a* is not expressed at levels high enough to be detected by northern blot analysis, and *DNMT3b* is expressed at equivalent levels in *DNMT1*^{+/+} and *DNMT1*^{-/-} cells (see Supplementary Information). Whether our findings reflect the different species (mouse versus human), cell types (germline versus somatic) or neoplastic derivation (normal versus cancer) represented by ES and HCT116 cells is not known; but this represents a promising area for future research. Our results also provide further evidence that genomic methylation is a highly site-specific process in human cells, with different methyltransferase activities being responsible for specific patterns of methylation at distinct loci²⁴. Such specificity offers a solution to the long-standing observation that cancer cells can show hypermethylation of specific CpG islands in the face of overall demethylation. □

Methods

DNMT1 targeting construct

A genomic clone containing most of *DNMT1* was obtained by screening a human bacterial artificial chromosome (BAC) library (Research Genetics). A 4.3-kilobase (kb) *HindIII*/*SpeI* fragment upstream of exon 3 and a 3.0-kb *EcoRI*/*NcoI* fragment downstream of exon 5 isolated from this BAC were incorporated on either side of a promoterless hygromycin resistance gene flanked by *loxP* sites. This strategy allowed Cre-mediated excision of the drug marker after targeting of the first allele and thus the use of the same targeting construct to delete the second allele. A PCR screen for targeted alleles was developed as described⁷. Further details about the targeting vector and screening PCR primers are available from the authors upon request.

Cell culture, transfection and screening for recombinants

HCT116 and HT29 cells (American Type Culture Collection) were cultured as described⁸. HCT116 cells were transfected with *HindIII*-linearized targeting vector using LipofectAMINE (Life Technologies) and selected in growth medium supplemented with 0.1 mg ml⁻¹ hygromycin (Calbiochem). Genomic DNA was prepared from drug-resistant clones using the QiaAmp 96 Spin Blood Kit (Qiagen) and analysed by PCR. Homologous recombinants identified by PCR were confirmed by Southern blot (Fig. 1). A *DNMT1*^{-/-} clone was infected with an adenovirus expressing Cre recombinase (refs 7, 25; and T.C.H. and B.V., unpublished data) to yield hygromycin-sensitive clones. One such clone was subjected to a second round of targeting, and three independent clones with both alleles of *DNMT1* disrupted were used for further experiments.

Western blot analysis and DNA methyltransferase assays

Freeze-thaw lysates of log-phase cells were used. Western blot analyses with antibodies against DNMT1 or Topoisomerase I (TopoGEN) were carried out with 200 µg of protein as described²⁶. Immunocomplexes were detected with a horseradish-peroxidase-conjugated goat anti-human secondary antibody (Pierce), and visualized using enhanced chemiluminescence (Amersham). DNA methyltransferase enzyme activity was measured by incubating 10 µg of protein lysate with 3 µCi of S-adenosyl-L-[methyl-³H]methionine (Amersham TK865) and 0.5 µg poly(dI-dC)-(dI-dC) (Pharmacia) or 2.5 µg oligonucleotide 5'-CCAGCCGCGCCGACCGACCGACCGCCGCGC-3' (methylated cytosines are underlined (Cruachem)) for 120 min at 37 °C. Reactions were stopped, purified by spin column (BioMax) and precipitated in ethanol. Resuspended nucleic acids were spotted on GF/C filter discs (Whatman) for liquid scintillation counting. Parallel reactions were carried out without DNA and were subtracted from the experimental values. Results are expressed as the mean disintegrations per minute (d.p.m.) per microgram of protein.

High performance liquid chromatography

Roughly 50 µg of RNA-free DNA was quantitatively digested with nuclease P1 (Roche Molecular Biochemicals) and calf intestinal alkaline phosphatase (Sigma) as described¹². Samples were separated on a reversed-phase column (Supelcosil LC-18 DB) as described¹³, monitoring absorbances at 275 nm and 285 nm. Peak assignments were confirmed using deoxyribonucleoside standards (Sigma). 5-Methylcytosine content was expressed as a percentage of the total cytosine pool, using peak areas after correction for extinction coefficients. For each clone, at least three separate aliquots of DNA were digested and chromatographed.

5-Aza-dC treatment, methylation-sensitive Southern blotting, and sodium bisulphite DNA sequencing

Growth medium supplemented with 5 µM 5-aza-dC (Sigma) was added to log-phase cells every 24 h over a 72-h period before isolating DNA. Genomic DNA from treated and

untreated cells was digested with *Bst*BI, *Hpa*II or *Msp*I (New England Biolabs), and Southern blotted using standard methods. Hybridization was performed with probes derived from aliphoid satellite templates from chromosomes 3, 4, 7, 10, 12, 17, 18, 22 and X (ref. 10); satellite 2 and 3 templates from chromosomes 1, 9, 15 and 16 (ref. 27); and *Alu*-I and *Alu*-Sx families²⁸. Genomic DNA from treated and untreated cells was modified by bisulphite treatment as described²⁹. *E-cadherin* *Alu* repeats were amplified with primers 5'-TAGGTAGTTAGAGAGAGAGTGTAGTGG-3' and 5'-CGTTTAAATAAATAAACTCACTCTTTC-3'. The CpG islands of *p16*^{INK4a} (ref. 9) and *TIMP-3* (ref. 30) were amplified as described. Polymerase chain reaction products were gel purified and ligated into pCR2.1-TOPO (Invitrogen). The drug sensitivity assay was performed as described¹⁹, using 24-h treatments with 0, 50, 100, 200, 500 and 1000 nM 5-aza-dC. Colonies were counted after two weeks of growth in drug-free media, and each measurement was made in triplicate.

Analysis of *p16*^{INK4a} and β-galactosidase expression

The region of *p16*^{INK4a} containing the frameshift mutation was amplified from complementary DNA using the primers 5'-TGGAGCCTTCGGCTGACT-3' and 5'-CTATGCGGGCATGGTTACTG-3' and from genomic DNA using the primers 5'-TGGAGCCTTCGGCTGACT-3' and 5'-CAATCCCTGCAAACTTCG-3'. All PCR products were gel purified and sequenced with the primer 5'-CTGGATCGGCTCCGAC-3' using Thermosequenase (Amersham). Exponentially growing cells were infected with the retrovirus Gal/Neo, which expresses the β-galactosidase reporter driven by a retroviral LTR and the G418 resistance gene under the control of the simian virus 40 early promoter, and resultant clones were analysed as described⁸.

Received 6 September 1999; accepted 10 February 2000.

- Baylin, S. B., Herman, J. G., Graff, J. R., Vertino, P. M. & Issa, J. P. Alterations in DNA methylation: a fundamental aspect of neoplasia. *Adv. Cancer Res.* **72**, 141–196 (1998).
- Jones, P. A. & Laird, P. W. Cancer epigenetics comes of age. *Nature Genet.* **21**, 163–167 (1999).
- Schmutte, C., Yang, A. S., Nguyen, T. T., Beart, R. W. & Jones, P. A. Mechanisms for the involvement of DNA methylation in colon carcinogenesis. *Cancer Res.* **56**, 2375–2381 (1996).
- Eads, C. A. *et al.* CpG island hypermethylation in human colorectal tumors is not associated with DNA methyltransferase overexpression. *Cancer Res.* **59**, 2302–2306 (1999).
- Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
- Hsu, D. W. *et al.* Two major forms of DNA (cytosine-5) methyltransferase in human somatic tissues. *Proc. Natl Acad. Sci. USA* **96**, 9751–9756 (1999).
- Chan, T. A., Hermeking, H., Lengauer, C., Kinzler, K. W. & Vogelstein, B. 14-3-3σ is required to prevent mitotic catastrophe after DNA damage. *Nature* **401**, 616–620 (1999).
- Lengauer, C., Kinzler, K. W. & Vogelstein, B. DNA methylation and genetic instability in colorectal cancer cells. *Proc. Natl Acad. Sci. USA* **94**, 2545–2550 (1997).
- Myohanen, S. K., Baylin, S. B. & Herman, J. G. Hypermethylation can selectively silence individual *p16*^{INK4A} alleles in neoplasia. *Cancer Res.* **58**, 591–593 (1998).
- Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instability in colorectal cancers. *Nature* **386**, 623–627 (1997).
- Graff, J. R., Herman, J. G., Myohanen, S., Baylin, S. B. & Vertino, P. M. Mapping patterns of CpG island methylation in normal and neoplastic cells implicates both upstream and downstream regions in *de novo* methylation. *J. Biol. Chem.* **272**, 22322–22329 (1997).
- Kuo, K. C., McCune, R. A., Gehrke, C. W., Midgett, R. & Ehrlich, M. Quantitative reversed-phase high performance liquid chromatographic determination of major and modified deoxyribonucleosides in DNA. *Nucleic Acids Res.* **8**, 4763–4776 (1980).
- Feinberg, A. P., Gehrke, C. W., Kuo, K. C. & Ehrlich, M. Reduced genomic 5-methylcytosine content in human colonic neoplasia. *Cancer Res.* **48**, 1159–1161 (1988).
- Frommer, M. *et al.* A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl Acad. Sci. USA* **89**, 1827–1831 (1992).
- Okamoto, A. *et al.* Mutations and altered expression of *p16*^{INK4} in human cancer. *Proc. Natl Acad. Sci. USA* **91**, 11045–11049 (1994).
- Bachman, K. E. *et al.* Methylation-associated silencing of the tissue inhibitor of metalloproteinase-3 gene suggests a suppressor role in kidney, brain, and other human cancers. *Cancer Res.* **59**, 798–802 (1999).
- Santi, D. V., Garrett, C. E. & Barr, P. J. On the mechanism of inhibition of DNA-cytosine methyltransferases by cytosine analogs. *Cell* **33**, 9–10 (1983).
- Santi, D. V., Normant, A. & Garrett, C. E. Covalent bond formation between a DNA-cytosine methyltransferase and DNA containing 5-azacytosine. *Proc. Natl Acad. Sci. USA* **81**, 6993–6997 (1984).
- Juttermann, R., Li, E. & Jaenisch, R. Toxicity of 5-aza-2'-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. *Proc. Natl Acad. Sci. USA* **91**, 11797–11801 (1994).
- Okano, M., Xie, S. & Li, E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature Genet.* **19**, 219–220 (1998).
- Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell* **99**, 247–257 (1999).
- Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366**, 362–365 (1993).
- Lei, H. *et al.* *De novo* DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* **122**, 3195–3205 (1996).
- Xu, G. L. *et al.* Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402**, 187–191 (1999).
- He, T. C. *et al.* A simplified system for generating recombinant adenoviruses. *Proc. Natl Acad. Sci. USA* **95**, 2509–2514 (1998).
- Vertino, P. M., Yen, R. W., Gao, J. & Baylin, S. B. *De novo* methylation of CpG island sequences in human fibroblasts overexpressing DNA (cytosine-5)-methyltransferase. *Mol. Cell. Biol.* **16**, 4555–4565 (1996).

27. Jeanpierre, M. Human satellites 2 and 3. *Ann. Genet.* **37**, 163–171 (1994).
28. Toyota, M. *et al.* Identification of differentially methylated sequences in colorectal cancer by methylated CpG island amplification. *Cancer Res.* **59**, 2307–2312 (1999).
29. Herman, J. G., Graff, J. R., Myohanen, S., Nelkin, B. D. & Baylin, S. B. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc. Natl Acad. Sci. USA* **93**, 9821–9826 (1996).
30. Cameron, E. E., Bachman, K. E., Myohanen, S., Herman, J. G. & Baylin, S. B. Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. *Nature Genet.* **21**, 103–107 (1999).

Supplementary information is available from Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank S. R. Lee and S. G. Rhee for sharing their expertise in HPLC analysis. We thank T. Chan, M. Esteller, N. Watkins and other members of our laboratories for helpful discussions. This work was supported by the Clayton Fund and by the National Institutes of Health. K.E.S. is a Fellow of the American Cancer Society.

Correspondence and requests for materials should be addressed to K.E.S. (e-mail: kornels@welchlink.welch.jhu.edu).

Direct observation of dendritic actin filament networks nucleated by Arp2/3 complex and WASP/Scar proteins

Laurent Blanchoin*, Kurt J. Amann*, Henry N. Higgs, Jean-Baptiste Marchand, Donald A. Kaiser & Thomas D. Pollard

Structural Biology Laboratory, Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, California 92037, USA

* These authors contributed equally to this work

Most nucleated cells crawl about by extending a pseudopod that is driven by the polymerization of actin filaments in the cytoplasm behind the leading edge of the plasma membrane^{1,2}. These actin filaments are linked into a network by Y-branches, with the pointed end of each filament attached to the side of another filament and the rapidly growing barbed end facing forward³. Because Arp2/3 complex nucleates actin polymerization and links the pointed end to the side of another filament *in vitro*, a dendritic nucleation model has been proposed⁴ in which Arp2/3 complex initiates filaments from the sides of older filaments. Here we report, by using a light microscopy assay, many new features of the mechanism. Branching occurs during, rather than after, nucleation by Arp2/3 complex activated by the Wiskott–Aldrich syndrome protein (WASP) or Scar protein; capping protein and profilin act synergistically with Arp2/3 complex to favour branched nucleation; phosphate release from aged actin filaments favours dissociation of Arp2/3 complex from the pointed ends of filaments; and branches created by Arp2/3 complex are relatively rigid. These properties result in the automatic assembly of the branched actin network after activation by proteins of the WASP/Scar family and favour the selective disassembly of proximal regions of the network.

We used fluorescence microscopy of filaments stained with fluorescent phalloidin (Figs 1 and 2) to establish the structural pathway of actin filament nucleation by Arp2/3 complex and to test the influence of profilin, capping protein, and nucleotide that is bound to actin monomers or actin filaments on the structure of the network. We focused on actin, Arp2/3 complex, capping protein and profilin, because they are ubiquitous⁵, abundant, essential in genetic systems⁶ and sufficient (with the addition of an actin-depolymerizing factor (ADF) or cofilin depolymerizing protein)

to reconstitute *in vitro* assembly of actin filament comet tails that drive the motility of the intracellular bacteria, *Listeria* and *Shigella*⁷. Arp2/3 complex links the pointed ends of actin filaments to the sides of other filaments at an angle of 70° (ref. 4) and nucleates growth in the barbed direction when activated by the WA domains of WASP/Scar proteins⁶. Actin filaments potentiate nucleation by Arp2/3 complex that is activated by WASP/Scar proteins^{8,9}. Arp2/3 complex is localized to 70° branches in the actin filament network at the leading edge of motile cells¹⁰. Capping protein and profilin have been proposed to maintain a pool of actin/profilin that is able to elongate free barbed ends but not pointed ends¹¹. Profilin inhibits spontaneous actin nucleation more strongly than nucleation by Arp2/3 complex⁸. Capping protein also terminates the elongation of barbed ends¹² and efficiently nucleates actin filaments that grow in the pointed direction^{13–15}.

Light microscopy provides the first direct evidence that branching occurs during actin filament nucleation by Arp2/3 complex and WASP/Scar proteins (Fig. 1). Results are similar with highly purified Arp2/3 complex from *Acanthamoeba* and bovine thymus activated by WA domains from human WASP and Scar-1 (Table 1). On their own, actin monomers form nuclei spontaneously at a low rate (Fig. 1a), resulting in long and unbranched filaments (Fig. 1c, Table 1). Few branches form if WA-domain activated Arp2/3 complex is added after completion of spontaneous polymerization (Fig. 1d). Arp2/3 complex activated by WA domains accelerates the polymerization of Mg-ATP-actin (Fig. 1a), creates new filaments (Fig. 1b) and produces branches (Fig. 1e and f) depending on the concentrations of both components (Table 1). Branch length is inversely proportional to branch density because the concentration of actin subunits is limiting. The concentration of new barbed ends plateaus at high concentrations of Arp2/3 complex (Fig. 1b)⁹, and, because nucleation is slower than elongation, actin monomers are consumed before each Arp2/3 complex can form a new filament. At high concentrations of activated Arp2/3 complex, branching is so dense that individual filaments cannot be resolved (data not shown). This is reminiscent of the densely branched filaments that are observed by electron microscopy in lamellipodia of keratocytes³. The mean lengths of actin filaments visualized by light microscopy are close to the lengths calculated from kinetic data (Table 1).

We determined the polarity of the filaments in branched networks by polymerizing Mg-ATP-actin monomers in the presence of rhodamine-phalloidin and preformed filaments labelled with Alexa green-phalloidin (Fig. 1j). In the absence of Arp2/3 complex, red unbranched actin filaments grow much longer at the barbed ends than at the pointed ends of the green filament seeds (Fig. 1j, left). In the presence of Arp2/3 complex and WASP/Scar proteins, the branches face the barbed ends, creating a polarized network (Fig. 1j, centre and right), as has been observed in cells by electron microscopy³. Dendritic nucleation occurs from the sides of ADP-P_i and ADP filaments with similar efficiency (Fig. 1j, centre and right). ADP-actin filaments were polymerized from ADP-actin monomers with green-phalloidin. ADP-P_i-actin filaments were polymerized from ATP-actin in the presence of green-phalloidin, which inhibits phosphate release. These filaments were added to ATP-actin monomers, activated-Arp2/3 complex and red phalloidin to label the new filaments. With ADP-P_i-actin filaments, 70% of the branches grew from the initial filaments, 30% from the new ATP-actin filaments. With ADP-P_i-actin filaments, 50% of the branches grew from the initial filaments. In biochemical experiments, ADP-actin and ATP-actin filaments enhance nucleation by Arp2/3 complex to the same extent⁹.

Nucleotide bound to actin has two marked effects on branching that are probably related mechanistically. First, ADP-actin monomers polymerize but rarely form branches in the presence of high concentrations of activated Arp2/3 complex (Fig. 1g). Second, the branching density of Mg-ATP-actin that is polymerized in the

presence of activated Arp2/3 complex depends on when rhodamine-phalloidin is added to the sample. Branching is maximal if rhodamine-phalloidin is present during polymerization. If added 20 min later, fewer branches are observed (Fig. 1h) unless BeF_3 is included (Fig. 1i). Because phalloidin inhibits phosphate release from ADP- P_i -actin filaments¹⁶ and BeF_3 stabilizes a state similar to ADP- P_i (ref. 17), ATP-actin and ADP- P_i -actin subunits probably favour the formation and maintenance of branches. ATP hydrolysis and phosphate dissociation may thus provide an automatic timer to initiate the disassembly of the branched network (Fig. 3).

The combination of Arp2/3 complex, capping protein and profilin strongly favours branched nucleation (Fig. 2b and d; and Table 1), creating many more short branches than in the absence of capping protein (Fig. 2a and c). Polymerization of actin with capping protein produces many short filaments that are capped on their barbed ends^{13,15}. Inclusion of profilin with activated Arp2/3 complex and capping protein suppresses both nucleation by capping protein and pointed-end growth from capped filaments,

but allows branching nucleation by Arp2/3 complex. Capping protein rapidly terminates these new filaments. This preserves the subunit pool and produces densely branched networks (Fig. 2b and d). This is a striking example of the cooperation of the multiple components in this complicated system.

High-speed imaging of branched actin filaments undergoing thermal fluctuations (Fig. 2e–n) revealed that the Y-junctions created by Arp2/3 complex are stiff. The mean ($\pm$ s.d.) angle measured 0.5–1 μm from the Y-junction in 400 frames of several branches was $77^\circ (\pm 13^\circ)$ for bovine Arp2/3 complex and $71^\circ (\pm 10^\circ)$ for amoeba Arp2/3 complex, which corresponds to rotational spring constants of $7.6 \times 10^{-20} \text{ J rad}^{-2}$ and $13 \times 10^{-20} \text{ J rad}^{-2}$, respectively. The narrow angular distribution suggests that branches produced by Arp2/3 complex are sufficiently rigid to be compatible with the elastic Brownian ratchet model¹⁸ for growing actin filaments pushing forward a membrane or a bacterium.

Light microscopy is an ideal approach to study the formation and dynamics of the branched actin network that is generated by

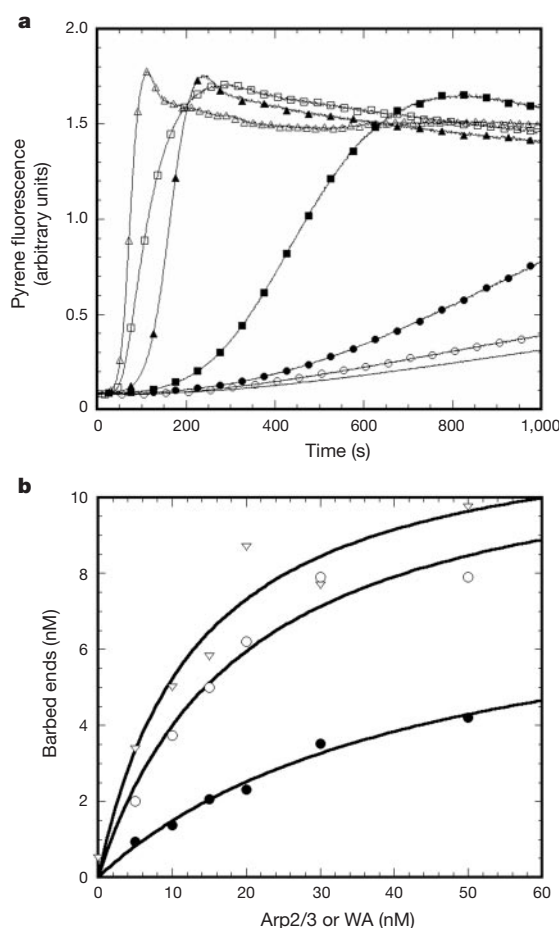
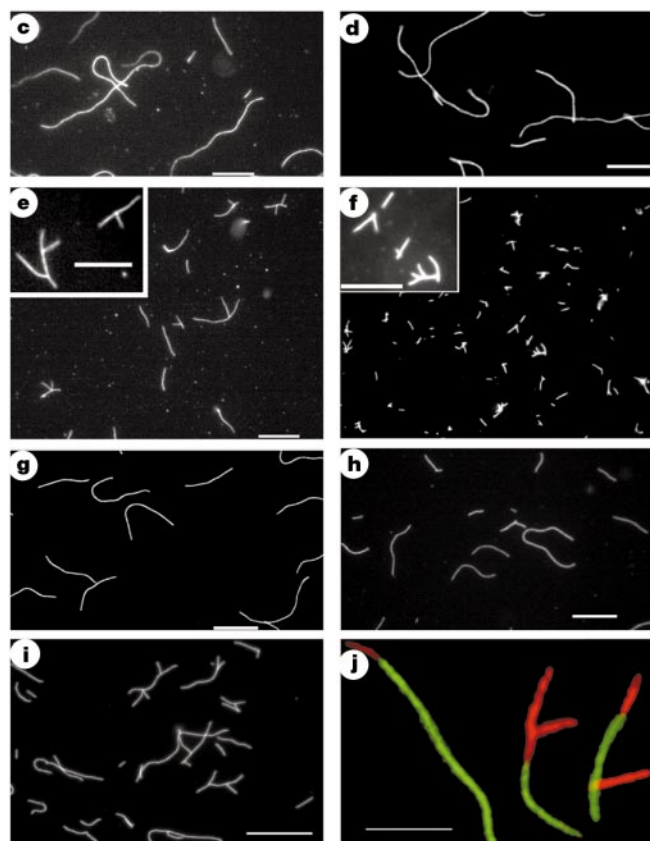


Figure 1 Comparison of the kinetics of actin polymerization in the presence of Arp2/3 complex and WASP WA domain with the structure of the products observed by light microscopy. **a,b**, Biochemical assays. Conditions: 4 μM actin (5% pyrene-labelled), 10 mM imidazole pH 7, 50 mM KCl, 1 mM MgCl_2 , 1 mM EGTA, 0.5 mM dithiothreitol, 0.1 mM CaCl_2 , 0.2 mM ATP and 3 mM $\text{Na}_2\text{S}_2\text{O}_8$, 22 $^\circ\text{C}$. **a**, Time course of polymerization monitored by pyrene fluorescence. Line with no symbols, no additions to actin. Filled symbols, experiments with 100 nM amoeba Arp2/3 complex and varying concentrations of human Scar WA: filled circles, no WA; filled squares, 5 nM WA; filled triangles, 50 nM WA. Open symbols, experiments with 100 nM human Scar WA and varying concentrations of Arp2/3 complex: open circles, no Arp2/3 complex; open squares, 5 nM Arp2/3 complex; open triangles, 50 nM Arp2/3 complex. **b**, Concentrations of barbed ends⁹. Open triangles, 100 nM WASP-WA with variable bovine Arp2/3 complex; open



circles, 100 nM Scar-WA with variable amoeba Arp2/3 complex; filled circles, 100 nM amoeba Arp2/3 complex with variable Scar-WA. **c–i**, Fluorescence micrographs of the products of actin polymerization as a function of Arp2/3 complex and WASP. **c**, Mg-ATP-actin alone; **d**, 20 nM of bovine Arp2/3 complex and 100 nM of WASP-WA added to the actin filament mixture after polymerization; **e**, Mg-ATP-actin polymerized with 15 nM bovine Arp2/3 complex and 100 nM WASP-WA; **f**, Mg-ATP-actin polymerized with 50 nM bovine Arp2/3 complex and 100 nM WASP-WA; **g**, Mg-ADP-actin polymerized with 50 nM bovine Arp2/3 complex and 400 nM WASP-WA; **h**, as **e** but phalloidin added after 20 min; **i**, as **h** but in the presence of 150 μM BeCl_2 and 5 mM NaF; **j**, as **e** but in presence of 300 nM polymerized actin previously labelled with Alexa green phalloidin. Scale bars, 10 μm for **c–i** and 2.5 μm for **j**.

dendritic nucleation in the presence of Arp2/3 complex. This approach cannot assess the components that are required for bacterial motility as can a reconstitution assay⁷, but the visualization of individual filaments provides more insights into the molecular mechanisms, which we summarize in a revised dendritic nucleation model (Fig. 3).

In our experiments and in cells, new filaments arise from a pool of Mg-ATP-actin monomers that are bound to profilin, which suppresses spontaneous nucleation. Existing filaments are mostly capped on their barbed ends by capping protein and on their pointed ends by Arp2/3 complex. Activation of Arp2/3 complex by a WASP/Scar protein initiates the assembly of a new filament (step 1). Branching is tightly coupled to nucleation (step 2). Arp2/3 complex may bind the side of a pre-existing filament before nucleation of a new filament, because preformed filaments stimulate nucleation⁸ and a lag occurs at the onset of polymerization of pure actin monomers, even with high concentrations of activated

Arp2/3 complex⁹. The pool of Mg-ATP-actin/profilin rapidly elongates the new barbed end (step 3), and the stiff, elastic filament anchored at its pointed end by a rigid branch pushes forward the plasma membrane (step 4). Elongation is transient, because growth is terminated by binding of capping protein (step 5). Profilin and capping protein enhance branching and limit the length of the new filaments. Profilin suppresses spontaneous nucleation more effectively than branching nucleation by Arp2/3 complex⁸, and capping protein terminates growth before the pool of subunits is depleted¹². Unexpectedly, the nucleotide that is bound to actin monomers and filaments influences dendritic nucleation and the stability of branches. The simplest interpretation of our observations is that ATP hydrolysis and phosphate release (step 6) weaken the interaction of the pointed end of the new filament with Arp2/3 complex (step 7) (R. D. Mullins, personal communication). This may be the first step in the disassembly of the new filament. ADF/cofilins promote this step by catalysing phosphate release¹⁹. These proteins

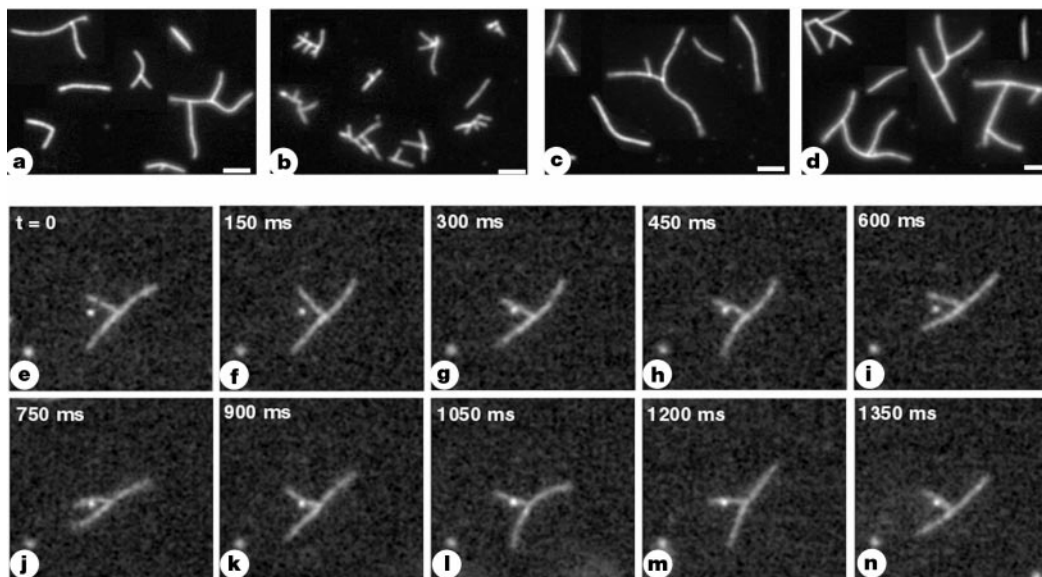


Figure 2 Regulation of branch length and observation of branch point rigidity. **a–d**, Effect of Arp2/3 complex and capping protein on actin polymerization. Polymerization conditions: 4 μ M actin monomers, 15 nM Arp2/3 complex, 100 nM WASP, 50 mM KCl, 1 mM MgCl₂, 1 mM EGTA, 0.5 mM DTT, 0.1 mM CaCl₂, 3 mM NaN₃, 10 mM imidazole pH 7.0 and 4 μ M rhodamine-phalloidin. **a,b**, Amoeba actin and 16 μ M amoeba profilin-I were polymerized alone (**a**), or with 100 nM amoeba capping protein (**b**). **c,d**, Muscle actin

and 8 μ M human profilin were polymerized alone (**c**), or with 60 nM amoeba capping protein (**d**). Scale bar, 2.5 μ m. **e–n**, Thermal fluctuations of branched actin filaments formed from 4 μ M muscle actin, 15 nM Arp2/3 complex, 100 nM WASP-WA and 4 μ M rhodamine-phalloidin and diluted into fluorescence buffer as in Fig. 1. Fluorescence micrographs were recorded at 15-ms intervals. Every tenth frame illustrates the stiffness of the branch point relative to the flexibility of the filament. Branch length 2 μ m.

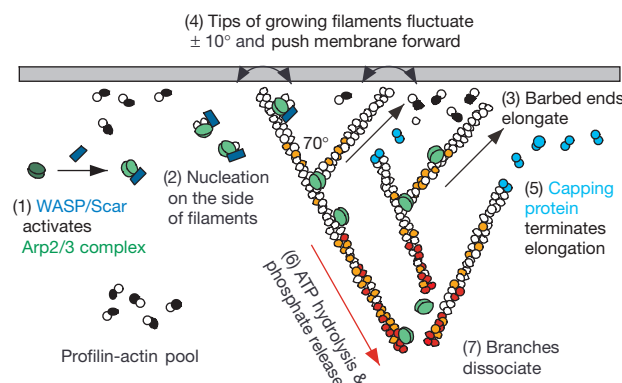


Figure 3 Model for the assembly and disassembly of the dendritic actin filament network at the leading edge of a cell. ATP-actin is shown in white, ADP-P-actin in orange, ADP-actin in red, and profilin in black.

Table 1 Quantitation of branching and filament length

Arp2/3 (nM)	WA (nM)	CP (nM)	Profilin (μM)	Biochemical length (μm)	Microscopic length (μm)	Branching (%)
Amoeba system: co-polymerization of Mg-ATP actin with rhodamine-phalloidin						
0	100	0	0		12.0 ± 10	0
5	100	0	0	6.5	7.8 ± 0.7	28
50	100	0	0	1.6	1.3 ± 0.5	97
15	100	0	16		2.4 ± 1.1	73
15	100	100	16		1.3 ± 0.6	91
Mammalian system: co-polymerization of Mg-ATP actin with rhodamine-phalloidin						
0	0	0	0	21	19.7 ± 13	0
10	100	0	0	2.6	6.6 ± 6	14
15	100	0	0		3.2 ± 2.2	56
50	100	0	0	1.4	2.3 ± 1.7	90
15	100	0	8		10.3 ± 10	8.3
15	100	60	8		5.8 ± 4.6	43
Addition of rhodamine-phalloidin after 20 min polymerization of Mg-ATP actin						
15	100	0	0		4.8 ± 4	12
Addition of rhodamine-phalloidin after 20 min polymerization of Mg-ATP actin with BeF ₃						
15	100	0	0		3.8 ± 4.2	42
Co-polymerization Mg-ATP actin with rhodamine-phalloidin; addition of Arp2/3 complex after 20 min						
15	100	0	0		15.0 ± 9.8	3
Co-polymerization of Mg-ADP actin with rhodamine-phalloidin						
60	400	0	0		12.6 ± 9.0	8

Protein sources: Amoeba actin was tested as indicated with amoeba Arp2/3 complex, human Scar-WA and amoeba profilin-1; and rabbit skeletal muscle actin was tested as indicated with bovine Arp2/3 complex, human WASp-WA and human profilin-I. CP, amoeba capping protein. Biochemical lengths are calculated as described⁹. Microscopic lengths are given with standard deviations. At least 100 individual filaments from 2 to 5 different experiments were observed for all length and branching measurements. Conditions: 4 μM actin with the indicated nucleotide, 4 μM rhodamine-phalloidin, 50 mM KCl, 1 mM MgCl₂, 1 mM EGTA, 10 mM imidazole pH 7.0, 1 mM ATP or ADP as indicated, 22 °C, for 20 min before dilution into fluorescence buffer and preparation of slides.

also sever ADP-actin filaments²⁰ and dissociate ADP-actin subunits from filament ends²¹. Further investigation of each of these steps will bring us closer to understanding the mechanism of force generation by actin polymerization during cell motility. □

Methods

Protein purification

Actin was purified from rabbit skeletal muscle acetone powder²² or from *Acanthamoeba*²³, Arp2/3 complex from bovine thymus⁹ or from *Acanthamoeba*²⁴, capping protein from *Acanthamoeba*²⁴, profilin from *Acanthamoeba*²⁶. Human profilin was bacterially expressed and purified²⁷. Actin was labelled on Cys 374 to a stoichiometry of 0.9–1.0 with pyrene iodoacetamide^{23,28}. Human WASP-WA domain was expressed and purified as described⁹. Human Scar-WA domain was bacterially expressed and purified by gel filtration (G50) followed by DEAE chromatography.

Polymerization

Actin was polymerized by addition of a 1:9 dilution of 10× KMEI (500 mM KCl, 10 mM MgCl₂, 10 mM EGTA, 100 mM imidazole, pH 7). Actin polymerization in presence of Arp2/3 complex and WASP-WA domain has been described⁹.

Microscopy

After 20 min of polymerization at 22 °C in presence of rhodamine-phalloidin or Alexa green phalloidin, actin was diluted to a final concentration of 10 nM in fluorescence buffer containing 50 mM KCl, 1 mM MgCl₂, 100 mM dithiothreitol, 20 μg ml⁻¹ catalase, 100 μg ml⁻¹ glucose oxidase, 3 mg ml⁻¹ glucose, 0.5% methylcellulose, 10 mM imidazole pH 7.0. Two microlitres were applied to coverslips coated in 0.1% nitrocellulose and fluorescence viewed with an Olympus IX-70 microscope using a mercury illumination source. Images were collected using a Hamamatsu digital camera. Filament lengths, branch number and angles were measured manually using Metamorph software. The rotational spring constant of branches was calculated from the equation, κ(θ)² = kT, based on the equipartition principle, where κ is the rotational spring constant, θ is the standard deviation of the angle (in radians) and kT is the thermal energy (4 × 10⁻²¹ J).

Received 30 December 1999; accepted 15 February 2000.

1. Wang, Y. Exchange of actin subunits at the leading edge of living fibroblasts: possible role of treadmilling. *J. Cell Biol.* **101**, 597–602 (1985).
2. Theriot, J. A. & Mitchison, T. J. Actin microfilament dynamics in locomoting cells. *Nature* **352**, 126–131 (1991).
3. Svitkina, T. M., Verkhovskiy, A. B., McQuade, K. M. & Borisy, G. G. Analysis of the actin-myosin II system in fish epidermal keratocytes: mechanism of cell body translocation. *J. Cell Biol.* **139**, 397–415 (1997).

4. Mullins, R. D., Heuser, J. A. & Pollard, T. D. The interaction of Arp2/3 complex with actin: nucleation, high-affinity pointed end capping, and formation of branching networks of filaments. *Proc. Natl Acad. Sci. USA* **95**, 6181–6186 (1998).
5. Pollard, T. D. in *Guidebook to the Cytoskeletal and Motor Proteins* 2nd edn (eds Kreis, T. & Vale, R.) 3–11 (1999).
6. Higgs, H. N. & Pollard, T. D. Regulation of actin polymerization by Arp2/3 complex and WASP/Scar proteins. *J. Biol. Chem.* **274**, 32531–32534 (1999).
7. Loisel, T. P., Boujemaa, R., Pantaloni, D. & Carlier, M. F. Reconstitution of actin-based motility of *Listeria* and *Shigella* using pure proteins. *Nature* **401**, 613–616 (1999).
8. Machesky, L. M. *et al.* Scar, A WASP-related protein, activates nucleation of actin filaments by the Arp2/3 complex. *Proc. Natl Acad. Sci. USA* **96**, 3739–3744 (1999).
9. Higgs, H. N., Blanchoin, L. & Pollard, T. D. Influence of the Wiskott–Aldrich syndrome protein (WASP) C terminus and Arp2/3 complex on actin polymerization. *Biochemistry* **38**, 15212–15222 (1999).
10. Svitkina, T. M. & Borisy, G. G. Arp2/3 complex and actin depolymerizing factor/cofilin in dendritic organization and treadmilling of actin filament array in lamellipodia. *J. Cell Biol.* **145**, 1009–1026 (1999).
11. Pollard, T. D. & Cooper, J. A. Quantitative analysis of the effect of *Acanthamoeba* profilin on actin filament nucleation and elongation. *Biochemistry* **23**, 6631–6641 (1984).
12. Schafer, D. A., Jennings, P. B. & Cooper, J. A. Dynamics of capping protein and actin assembly *in vitro*: uncapping barbed ends by polyphosphoinositides. *J. Cell Biol.* **135**, 169–179 (1996).
13. Cooper, J. A. & Pollard, T. D. Effects of capping protein on the kinetics of actin polymerization. *Biochemistry* **24**, 793–799 (1985).
14. Caldwell, J. E., Heiss, S. G., Mermall, V. & Cooper, J. A. Effects of CapZ, an actin capping protein of muscle, on the polymerization of actin. *Biochemistry* **28**, 8506–8514 (1989).
15. Xu, J., Casella, J. F. & Pollard, T. D. Effect of capping protein, CapZ, on the length of actin filaments and mechanical properties of actin filament networks. *Cell Motil. Cytoskeleton* **42**, 73–81 (1999).
16. Dancker, P. & Hess, L. Phalloidin reduces the release of inorganic phosphate during actin polymerization. *Biochim. Biophys. Acta* **1035**, 197–200 (1990).
17. Combeau, C. & Carlier, M. F. Characterization of the aluminum and beryllium fluoride species bound to F-actin and microtubules at the site of the g-phosphate of the nucleotide. *J. Biol. Chem.* **264**, 19017–19021 (1989).
18. Mogilner, A. & Oster, G. Cell motility driven by actin polymerization. *Biophys. J.* **71**, 303–345 (1996).
19. Blanchoin, L. & Pollard, T. D. Mechanism of interaction of *Acanthamoeba* actophorin (ADF/cofilin) with actin filaments. *J. Biol. Chem.* **274**, 15538–15546 (1999).
20. Maciver, S. K., Zot, H. G. & Pollard, T. D. Characterization of actin filament severing by actophorin from *Acanthamoeba castellanii*. *J. Cell Biol.* **115**, 1611–1620 (1991).
21. Carlier, M. F. *et al.* Actin depolymerizing factor (ADF/cofilin) enhances the rate of filament turnover: implication in actin-based motility. *J. Cell Biol.* **136**, 1307–1322 (1997).
22. Spudich, J. A. & Watt, S. The regulation of rabbit skeletal muscle contraction. Biochemical studies of the interaction of the tropomyosin–troponin complex with actin and the proteolytic fragments of myosin. *J. Biol. Chem.* **246**, 4866–4871 (1971).
23. Pollard, T. D. Polymerization of ADP-actin. *J. Cell Biol.* **99**, 769–777 (1984).

24. Machesky, L. M., Atkinson, S. J., Ampe, C., Vandekerckhove, J. & Pollard, T. D. Purification of a cortical complex containing two unconventional actins from *Acanthamoeba* by affinity chromatography on profilin agarose. *J. Cell Biol.* **127**, 107–115 (1994).
25. Cooper, J. A., Blum, J. D. & Pollard, T. D. *Acanthamoeba castellanii* capping protein: properties, mechanism of action, immunologic cross-reactivity, and localization. *J. Cell Biol.* **99**, 217–225 (1984).
26. Kaiser, D. A., Goldschmidt-Clermont, P. J., Levine, B. A. & Pollard, T. D. Characterization of renatured profilin purified by urea elution from poly-L-proline agarose columns. *Cell Motil.* **14**, 251–262 (1989).
27. Almo, S. C., Pollard, T. D., Way, M. & Lattman, E. E. Purification, characterization and crystallization of *Acanthamoeba* profilin expressed in *Escherichia coli*. *J. Mol. Biol.* **236**, 950–952 (1994).
28. Kouyama, T. & Mihashi, K. Fluorimetry study of N-(1-pyrenyl) iodoacetamide-labelled F-actin. Local structural change of actin protomer both on polymerization and on binding of heavy meromyosin. *Eur. J. Biochem.* **114**, 33–38 (1981).

Acknowledgements

This work was supported by an NIH research grant to T.D.P., fellowships from National Research Service Award to K.J.A. and H.N.H. and from the Association pour la Recherche contre le Cancer to J.B.M. We thank J. Howard for advice on calculating the spring constant.

Correspondence and requests for materials should be addressed to T.D.P. (e-mail: pollard@salk.edu).

Low fidelity DNA synthesis by human DNA polymerase- η

Toshiro Matsuda*, Katarzyna Bebenek*, Chikahide Masutani†, Fumio Hanaoka†‡ & Thomas A. Kunkel*§

* Laboratory of Molecular Genetics and § Laboratory of Structural Biology, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, USA

† Institute for Molecular and Cellular Biology, Osaka University and CREST, Japan Science and Technology Corporation, 1-3 Yamada-oka, Suita, Osaka 565-0871, Japan

‡ RIKEN, Wako-shi, Saitama 351-0198, Japan

A superfamily of DNA polymerases that bypass lesions in DNA has been described^{1–4}. Some family members are described as error-prone because mutations that inactivate the polymerase reduce damage-induced mutagenesis. In contrast, mutations in the skin cancer susceptibility gene *XPV*^{5,6}, which encodes DNA polymerase (pol)- η , lead to increased ultraviolet-induced mutagenesis^{7–11}. This, and the fact that pol- η primarily inserts adenines during efficient bypass of thymine–thymine dimers *in vitro*^{8,12,13}, has led to the description of pol- η as error-free. However, here we show that human pol- η copies undamaged DNA with much lower fidelity than any other template-dependent DNA polymerase studied. Pol- η lacks an intrinsic proofreading exonuclease activity and, depending on the mismatch, makes one base substitution error for every 18 to 380 nucleotides synthesized. This very low fidelity indicates a relaxed requirement for correct base pairing geometry and indicates that the function of pol- η may be tightly controlled to prevent potentially mutagenic DNA synthesis.

The ability of pol- η to bypass efficiently a DNA adduct that usually blocks synthesis by other polymerases indicates that pol- η may have relaxed discrimination ability. To test this hypothesis, we examined the fidelity of human pol- η during copying of undamaged DNA. Recombinant pol- η containing a carboxy-terminal hexahistidine tag⁵ was first examined for the presence of exonuclease activities using a 5'-end labelled template primer containing a 3'-terminal A-dGMP mismatch¹⁴. Neither 5' $\rightarrow$ 3' nor 3' $\rightarrow$ 5' exonuclease activities were detected (Fig. 1a, lane 2). Thus, pol- η lacks an intrinsic exonuclease activity that can proof-read replication errors. We then investigated the fidelity of pol- η during synthesis to fill a five-nucleotide gap with an undamaged

DNA template containing a TGA nonsense codon in the *LacZ* gene in M13mp2 DNA¹⁵. This substrate encodes a colourless M13 plaque phenotype, and errors are scored as blue revertant plaques when the DNA products are introduced into an α -complementation strain of *Escherichia coli* and plated on indicator plates.

Synthesis by pol- η generated gap-filled products (not shown, but see ref. 15 and below) whose revertant frequency was 8,400-fold higher than that of control DNA that had not been copied *in vitro* (Table 1). DNA sequence analysis revealed that 8 of 12 revertants contained a single base substitution and that the other 4 contained 2 substitutions among the 5 nucleotides copied. Both the high reversion frequency and the number of revertants containing multiple substitutions indicate that pol- η has low fidelity. The base substitution error rate calculated from these data is $3,100 \times 10^{-5}$ (1 error in 32 bases, 1/32). This rate is 24-fold higher than for human DNA polymerase- β (Table 1), the least accurate among wild-type DNA polymerases that have previously been studied.

To determine whether such low fidelity is site-specific or a general feature of synthesis by pol- η , we used a 407-nucleotide gap that permits detection of hundreds of different substitution, addition, deletion and complex errors generated when copying a 275-nucleotide sequence of the wild-type *LacZ* α -complementation gene¹⁶. Here, errors are scored as light-blue or colourless mutant plaques and the error specificity is defined by sequence analysis of independent *lacZ* mutants. Analysis of the DNA products of gap-filling synthesis revealed that pol- η filled the 407-nucleotide gap (Fig. 1b, lane 2). These products yielded a *lacZ* mutant frequency of 32%, a value that is unprecedented in comparison with the 0.1 to 3% values obtained with other wild-type, template-dependent DNA polymerases (Table 1). Similar values were observed for synthesis using 10- and 100-fold lower dNTP concentrations or for reactions containing 2- or 4-fold less pol- η . The average mutant frequency for six independent determinations was 34% (Table 1).

In contrast to the predominantly light-blue phenotypes of *lacZ* mutants generated by most other DNA polymerases using this assay, more than 95% of the pol- η -dependent *lacZ* mutants had a colourless rather than light-blue plaque phenotype. Sequence analysis of 24 *lacZ* mutants revealed that they contained from 3 to 22 changes per mutant. A total of 232 single base substitutions, 10 tandem double-base substitutions, 13 additions of 1–3 nucleotides and 32 deletions of 1–56 nucleotides were observed. The recovery of 232 single base substitutions among only 6,600 total nucleotides analysed (24 mutants $\times$ 275 nucleotides) yields an average base

Table 1 Fidelity of human pol- η

DNA polymerase	Polymerase family	3'-exo activity	Mutant frequency* ($\times 10^{-4}$)	Average base substitution rate ($\times 10^{-5}$)†
TGA codon reversion assay			0.05 (control‡)	
Pol- η	Rad30	no	420	3,100
Pol- β	Pol X	no	26	130§
Forward mutation assay			6 (control‡)	
Pol- η	Rad30	no	3,400	3,500
Pol- β	Pol X	no		67
Pol- α	Pol α	no		16
HIV-1 RT	RT	no		6
<i>E. coli</i> Kf pol	Pol I	no		2
Pol- δ	Pol α	yes		~ 1
Pol- ϵ	Pol α	yes		≤ 1
Pol- γ	Pol I	yes		≤ 1

* The reversion frequency is the average of two determinations; the forward mutant frequency is the average for six experiments where reaction conditions were varied and yielded the following mutant frequencies: 5 mM MgCl₂, 10 μ M dNTPs: 32%; 5 mM MgCl₂, 100 μ M dNTPs: 37%; 10 mM MgCl₂, 1 mM dNTPs: (18 nM pol- η) 39%, (36 nM pol- η) 37%, (72 nM pol- η) 28 and 32%. In each case, about 1,000 total plaques were scored.

† Error rates for the other DNA polymerases^{14,15,27,28} are averages, which vary depending on the mismatch and its location.

‡ Control values are for DNA substrates not subjected to copying *in vitro*.

§ From ref. 16.

|| The value for exonuclease-deficient Klenow fragment¹⁴.

substitution error rate of 3.5% (Table 1), which is one error for every 28 nucleotides polymerized. We recovered errors consistent with formation of 11 of the 12 possible mispairs (Table 2). Among the 12 mispairs, error rates ranged from 1/18 for a T-dGMP mispair to greater than 1/1900 for a C-dCMP mispair. These values represent the average rates for 68, 67, 61 and 79 template positions for A, T, G and C, respectively (Table 2). The overall average rate of one in 28 is more than 3,000-fold higher than that of eukaryotic replicative DNA polymerases δ , ϵ , and γ , for results obtained using the same assay. The lower fidelity of pol- η is not merely due to the lack of a proofreading exonuclease activity, as pol- η is 50- to 1,700-fold less accurate than four other exonuclease-deficient DNA polymerases measured with the same assay (Table 1). Further evidence that pol- η has unusually low fidelity when copying undamaged DNA is the recovery of 10 tandem double base substitutions among only 24 mutants. Such mutations require three consecutive rare events—an initial misinsertion, a second misinsertion using a mismatched template primer, and correct extension from a doubly mismatched template primer. The data in Table 1 reveal that pol- η is much less accurate than are representatives of each of the four other DNA polymerase families.

We next examined the infidelity of pol- η with an independent experimental approach, steady-state kinetic analysis (Table 3). Pol- η inserted correct dAMP opposite template T and correct dCMP opposite template G with catalytic efficiencies (k_{cat}/K_m) that are similar to that observed with HIV-1 reverse transcriptase (RT) (Table 3). Thus, pol- η has robust catalytic efficiency for formation of correct T-A and G-C base pairs. However, it also has low nucleotide selectivity, as shown by the high rate of misinsertion of incorrect dGMP or dCMP opposite template T, or of incorrect dGMP opposite template G (Table 3). Note that discrimination is low because of efficient insertion of incorrect nucleotides, rather than inefficient insertion of correct nucleotides. The misinsertion rates ($1/f_{ins}$) for these three mispairs at these two template locations are remarkably similar to the average error rates for the same

mispairs observed during gap-filling synthesis (Table 2). Moreover, the pol- η misinsertion rate for the T-dGMP mispair is much higher than that of HIV-1 RT (Table 3), again consistent with the results of gap-filling reactions (Table 1). Thus, an independent experimental approach confirms the low fidelity of pol- η . This may be a common property of pol- η regardless of source, as yeast pol- η also appears to have low nucleotide insertion fidelity with undamaged templates¹⁷.

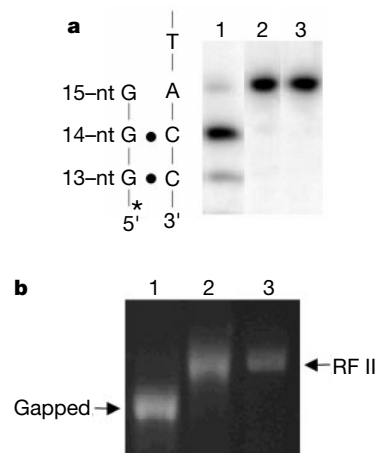


Figure 1 Exonuclease activities and gap-filling synthesis by pol- η . **a**, Exonuclease assay using Klenow fragment (lane 1), pol- η (lane 2) or no enzyme (lane 3). The total pixels in the lanes were similar, indicating no detectable 5'-exonuclease. The $n-1$ and $n-2$ products generated by pol- η indicate a specific exonuclease activity (% digestion per pmol per min) 80-fold lower than for Klenow fragment, which has modest proofreading capacity¹⁴. **b**, Products of gap-filling synthesis examined by agarose gel electrophoresis. The reaction included no enzyme (lane 1), pol- η (lane 2) or nicked, double-stranded (RFII) DNA standard (lane 3). nt, nucleotide. Asterisk, ³²P labelling.

Table 2 Base substitution error rates of pol- η

Base	Number* (total)†	Mutation from → to	Mismatch template-dNMP	Observed substitutions	Error rate
A	68 (1,632)	A → G	A-dCMP	17	1/96
		A → T	A-dAMP	26	1/63
		A → C	A-dGMP	21	1/78
T	67 (1,608)	T → C	T-dGMP	91	1/18
		T → A	T-dTMP	13	1/120
		T → G	T-dCMP	8	1/200
G	61 (1,464)	G → A	G-dTMP	21	1/70
		G → C	G-dGMP	4	1/370
		G → T	G-dAMP	7	1/210
C	79 (1,896)	C → T	C-dAMP	19	1/100
		C → G	C-dCMP	0	≤1/1,900
		C → A	C-dTMP	5	1/380

* The number of template bases within the 275-base LacZ target sequence.

† The total number of template bases in 24 sequenced mutants.

Table 3 Kinetic analysis of misinsertion by pol- η

Enzyme	Template nucleotide	dNTP	K_m (μ M)	k_{cat} (min^{-1})	(k_{cat}/K_m) ($\mu\text{M}^{-1} \times 10^3$)	$f_{ins} \times 10^{3*}$
Pol- η	T	dATP	31 ± 6	13 ± 3.2	420 ± 130	—
	T	dGTP	250 ± 84	5.5 ± 2.2	22 ± 10	52 (1/19)
	T	dCTP	2,900 ± 1,600	4.8 ± 0.7	1.6 ± 0.9	3.9 (1/260)
	G	dCTP	11 ± 1.9	8.5 ± 1.9	760 ± 210	—
	G	dGTP	450 ± 111	3.9 ± 0.4	8.7	11† (1/90)
	G	dATP	0.20 ± 0.05	0.16 ± 0.01	800	—
HIV-1 RT	T	dATP	0.20 ± 0.05	0.16 ± 0.01	800	—
	T	dGTP	140 ± 24	0.01 ± 0.0005	0.07	0.1 (1/10,000)

Reactions were performed as described in Methods and kinetic constants were derived as described²⁶.

* f_{ins} is defined as $(k_{cat}/K_m)_{incorrect}/(k_{cat}/K_m)_{correct}$. For ease of comparison with the error rates in Table 2, values for $1/f_{ins}$ are shown in parentheses. The results for HIV-1 RT are from ref. 26.

† Some of this insertion could be templated by the adjacent 5'-template C, through misalignment.

Several possibilities exist to reconcile the observation that pol- η has low fidelity but is also involved in a process that reduces ultraviolet-induced mutagenesis. Pol- η synthesis past lesions could be more accurate than copying of undamaged DNA, or fidelity may be enhanced by replication accessory proteins. Alternatively, pol- η misinsertions could be proofread by a separate exonuclease¹⁸ or removed by DNA mismatch repair¹⁹. The low fidelity of pol- η during copying of undamaged DNA and its ability to fill gaps typical of those generated during base and nucleotide excision repair suggest that controls may have evolved to limit synthesis by pol- η . The opportunity to generate errors *in vivo* may be limited by the need to form partnerships with specific replication accessory proteins^{1–3}, and pol- η may only synthesize enough nucleotides to generate a duplex template primer that can be elongated by other, more accurate DNA polymerases. Pol- η function may also be controlled at the level of messenger RNA synthesis²⁰ or stability, post-translational modification or protein stability. The latter possibility is consistent with the mutator phenotype observed upon disruption of the yeast *RAD6* gene²¹. *RAD6* is a ubiquitin conjugase and its inactivation could result in inefficient turnover of yeast pol- η , resulting in enhanced mutagenesis owing to inaccurate DNA synthesis. Indeed, precedents exist for enhanced mutagenesis resulting from an excess of a polymerase. Elevated spontaneous mutation rates have been observed upon overexpression of a mutator mutant of pol- β ²², the *E. coli* DinB polymerase²³ or the murine DINB1 polymerase⁴. *E. coli* DinB polymerase and DINB1 are members of the superfamily of DNA polymerases to which human pol- η belongs. Structural and functional studies of the other four DNA polymerase families reveal a strict requirement for correct Watson–Crick base pair geometry (reviewed in ref. 24). Those enzymes copy undamaged templates with higher fidelity than pol- η , and they have limited ability to bypass bulky adducts. The lower fidelity of pol- η (Table 2) implies that it may incorporate nucleotides that are primarily stabilized by base–base hydrogen bonding and/or stacking interactions and that it may not require strict shape complementarity to bypass bulk lesions efficiently. In fact, there may be an inverse relationship between lesion bypass efficiency and nucleotide selectivity. This hypothesis can be tested by examining the fidelity of other enzymes that bypass lesions, such as pol- ζ , which also bypasses thymine–thymine dimers but at a lower efficiency²⁵. □

Methods

Assay for exonuclease activities

Reactions (25 μ l) contained 40 mM Tris-HCl (pH 8.0), 10 mM MgCl₂, 10 mM DTT, 6.25 μ g BSA, 60 mM KCl, 2.5% glycerol, 2.8 pM 5'-³²P-labelled template primer DNA containing a 3'-terminal A-G mismatch¹⁴, and either 96 nM pol- η or 150 nM of exonuclease-proficient Klenow polymerase (Kf). Reactions were incubated at 37 °C for 5 min (Kf) or 30 min (pol- η) and examined by gel electrophoresis. The products were quantified by phosphorimager. A reduction in total signal intensity indicates the presence of 5'-nuclease activity. Products shorter than 15 nucleotides indicate the presence of 3'-exonuclease activity, with a product of 14 bp indicating excision of the terminal mismatch.

Gap-filling synthesis and fidelity measurements

The sources and descriptions of strains and reagents, and preparation of substrates were as described^{15,16}. Reactions (25 μ l) contained 40 mM Tris-HCl (pH 8.0), 10 mM DTT, 6.25 μ g BSA, 60 mM KCl, 2.5% glycerol, 10 mM MgCl₂, 1 mM dNTPs (forward assay) or 0.5 mM dNTPs (reversion assay), 1.4 nM M13mp2 DNA containing a 407-nucleotide gap (from nucleotides, –216 to +191 of the *LacZ* gene), and 72 nM pol- η (about 50:1 ratio of enzyme to DNA). Reactions were incubated at 37 °C for 1 h and terminated by adding EDTA to 15 mM. DNA products were processed for mutant frequencies as described^{15,16}. As a control, when pure heteroduplex DNA containing a single mismatch is introduced into the *E. coli* strain used to score polymerase errors, the percentage of M13 plaques having the phenotype of the (newly-synthesized) minus strand is typically about 60% (ref. 16). However, we have previously shown that minus-strand expression values decrease as the number of unpaired nucleotides in the minus strand increases. The high frequency of forward mutants observed for the pol- η products (Table 1) and the multiple mutations in each *LacZ* mutant indicate that all minus strand products of pol- η gap-filling that survived and were expressed in *E. coli* containing errors made by pol- η . Consistent with this, when the DNA from six dark blue M13 clones obtained from pol- η reactions were sequenced, no mutations were found.

Kinetic analysis of misinsertion

Reactions (25 μ l) were as for gap-filling, except for use of 200 nM 30-nucleotide template primed at a 1.2 to 1 molar ratio with a 5'-³²P-labelled 20 bp oligonucleotide complementary to nucleotides 88–107 or 89–108 of the *LacZ* gene, 2 nM pol- η , and correct dATP or dCTP at concentrations of 1 to 100 μ M or incorrect dGTP or dCTP at concentrations of 5 to 3,000 μ M. With one exception (one determination with the G-G mismatch), duplicate determinations were performed for each template-dNTP combination at seven different dNTP concentrations. Aliquots were removed at 2, 4, 6 and 8 min and the products separated by gel electrophoresis and quantified by phosphorimager. Kinetic constants were derived as described²⁶.

Received 6 December 1999; accepted 11 February 2000.

- Friedberg, E. C. & Gerlach, V. L. Novel DNA polymerases offer clues to the molecular basis of mutagenesis. *Cell* **98**, 413–416 (1999).
- Woodgate, R. A plethora of lesion-replicating DNA polymerases. *Genes Dev.* **13**, 2191–2195 (1999).
- Johnson, R. E., Washington, M. T., Prakash, S. & Prakash, L. Bridging the gap: A family of novel DNA polymerases that replicate faulty DNA. *Proc. Natl Acad. Sci. USA* **96**, 12224–12226 (1999).
- Ogi, T., Kato, T. Jr, Kato, T. & Ohmori, H. Mutation enhancement by DINB1, a mammalian homologue of the *Escherichia coli* mutagenesis protein DinB. *Genes Cells* **4**, 607–618 (1999).
- Masutani, C. *et al.* The XPD (xeroderma pigmentosum variant) gene encodes human DNA polymerase η . *Nature* **399**, 700–704 (1999).
- Johnson, R. E., Kondratieff, C. M., Prakash, S. & Prakash, L. hRAD30 mutations in the variant form of xeroderma pigmentosum. *Science* **285**, 263–265 (1999).
- Maher, V. M., Ouellette, L. M., Curren, R. D. & McCormick, J. J. Frequency of ultraviolet light-induced mutations is higher in xeroderma pigmentosum variant cells than in normal human cells. *Nature* **261**, 593–595 (1976).
- Myhr, B. C., Turnbull, D. & DiPaolo, J. A. Ultraviolet mutagenesis of normal and xeroderma pigmentosum variant human fibroblasts. *Mutat. Res.* **62**, 341–353 (1979).
- Wang, Y. C., Maher, V. M. & McCormick, J. J. Xeroderma pigmentosum variant cells are less likely than normal cells to incorporate dAMP opposite photoproducts during replication of UV-irradiated plasmids. *Proc. Natl Acad. Sci. USA* **88**, 7810–7814 (1991).
- Wang, Y. C., Maher, V. M., Mitchell, D. L. & McCormick, J. J. Evidence from mutation spectra that the UV hypermutability of xeroderma pigmentosum variant cell reflects abnormal, error-prone replication on a template containing photoproducts. *Mol. Cell. Biol.* **13**, 4276–4283 (1993).
- Waters, H. L., Seetharam, S., Seidman, M. M. & Kraemer, K. H. Ultraviolet hypermutability of a shuttle vector propagated in xeroderma pigmentosum variant cells. *J. Invest. Dermatol.* **101**, 744–748 (1993).
- Masutani, C. *et al.* Xeroderma pigmentosum variant (XP-V) correcting protein from HeLa cells has a thymine dimer bypass DNA polymerase activity. *EMBO J.* **18**, 3491–3501 (1999).
- Johnson, R. E., Prakash, S. & Prakash, L. Efficient bypass of a thymine-thymine dimer by yeast DNA polymerase, Pol η . *Science* **283**, 1001–1004 (1999).
- Roberts, J. D., Bebenek, K. & Kunkel, T. A. The accuracy of reverse transcriptase from HIV-1. *Science* **242**, 1171–1173 (1988).
- Osheroff, W. P. *et al.* The fidelity of DNA polymerase beta during distributive and processive DNA synthesis. *J. Biol. Chem.* **274**, 3642–3650 (1999).
- Bebenek, K. & Kunkel, T. A. Analyzing the fidelity of DNA polymerases. *Methods Enzymol.* **262**, 217–232 (1995).
- Washington, M. T., Johnson, R. E., Prakash, S. & Prakash, L. Fidelity and processivity of *Saccharomyces cerevisiae* DNA polymerase η . *J. Biol. Chem.* **274**, 36835–36838 (1999).
- Perrino, F. W. & Loeb, L. A. Differential extension of 3' mispairs is a major contribution to the high fidelity of calf thymus DNA polymerase- α . *J. Biol. Chem.* **264**, 2898–2905 (1989).
- Wang, H., Lawrence, C. W., Li, G. M. & Hays, J. B. Specific binding of human MSH2-MSH6 mismatch-repair protein heterodimers to DNA incorporating thymine- or uracil-containing UV light photoproducts opposite mismatched bases. *J. Biol. Chem.* **274**, 16894–16900 (1999).
- McDonald, J. P., Levine, A. S. & Woodgate, R. The *Saccharomyces cerevisiae* *RAD30* gene, a homologue of *Escherichia coli* *dinB* and *umuC*, is damage inducible and functions in a novel error-free postreplication repair mechanism. *Genetics* **147**, 1557–1568 (1997).
- Kang, X. L., Yadao, F., Gietz, R. D. & Kunz, B. A. Elimination of the yeast *RAD6* ubiquitin conjugase enhances base-pair transitions and G.C to T.A transversions as well as transposition of the Ty element: implications for the control of spontaneous mutation. *Genetics* **130**, 285–294 (1992).
- Clairmont, C. A. *et al.* The Tyr-265-to-Cys mutator mutant of DNA polymerase beta induces a mutator phenotype in mouse LN12 cells. *Proc. Natl Acad. Sci. USA* **96**, 9580–9585 (1999).
- Kim, S. R. *et al.* Multiple pathways for SOS-induced mutagenesis in *Escherichia coli*: an overexpression of *dinB/dinP* results in strongly enhancing mutagenesis in the absence of any exogenous treatment to damage DNA. *Proc. Natl Acad. Sci. USA* **94**, 13792–13797 (1997).
- Kunkel, T. A. & Bebenek, K. DNA replication fidelity. *Annu. Rev. Biochem.* **69**, (in the press).
- Nelson, J. R., Lawrence, C. W. & Hinkle, D. C. Thymine-thymine dimer bypass by yeast DNA polymerase ζ . *Science* **272**, 1646–1649 (1996).
- Lewis, D. A. *et al.* Uniquely altered DNA replication fidelity conferred by an amino acid change in the nucleotide binding pocket of human immunodeficiency virus type 1 reverse transcriptase. *J. Biol. Chem.* **274**, 32924–32930 (1999).
- Bebenek, K., Joyce, C. M., Fitzgerald, M. P. & Kunkel, T. A. The fidelity of DNA synthesis catalyzed by derivatives of *Escherichia coli* DNA polymerase I. *J. Biol. Chem.* **265**, 13878–13887 (1990).
- Roberts, J. D. & Kunkel, T. A. in *DNA Replication in Eukaryotic Cells: Concepts, Enzymes and Systems* (ed. Pampphilis, M. D.) 217–247 (Cold Spring Harbor Laboratory Press, New York, 1996).

Acknowledgements

We thank W. A. Beard and Y. Pavlov for critical evaluation of the manuscript.

Correspondence and requests for materials should be addressed to T.A.K. (e-mail: kunkel@niehs.nih.gov).

Roles of *E. coli* DNA polymerases IV and V in lesion-targeted and untargeted SOS mutagenesis

Mengjia Tang*, Phuong Pham*, Xuan Shen*, John-Stephen Taylor†, Mike O'Donnell‡, Roger Woodgate§ & Myron F. Goodman*

* Department of Biological Sciences and Chemistry, University of Southern California, University Park, Los Angeles, California 90089-1340, USA

† Department of Chemistry, Washington University, St. Louis, Missouri 63130, USA

‡ Rockefeller University and Howard Hughes Medical Institute, New York, New York 10021, USA

§ Section on DNA Replication, Repair and Mutagenesis, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892-2725, USA

The expression of the *Escherichia coli* DNA polymerases pol V (UmuD'2C complex)^{1,2} and pol IV (DinB)³ increases in response to DNA damage⁴. The induction of pol V is accompanied by a substantial increase in mutations targeted at DNA template lesions in a process called SOS-induced error-prone repair⁴. Here we show that the common DNA template lesions, TT (6–4) photoproducts, TT *cis-syn* photodimers and abasic sites, are efficiently bypassed within 30 seconds by pol V in the presence of activated RecA protein (RecA*), single-stranded binding protein (SSB) and pol III's processivity β , γ -complex. There is no detectable bypass by either pol IV or pol III on this time scale. A mutagenic 'signature' for pol V is its incorporation of guanine opposite the 3'-thymine of a TT (6–4) photoproduct, in agreement with mutational spectra. In contrast, pol III and pol IV incorporate adenine almost exclusively. When copying undamaged DNA, pol V exhibits low fidelity with error rates of around 10^{-3} to 10^{-4} , with pol IV being 5- to 10-fold more accurate. The effects of RecA protein on pol V, and β , γ -complex on pol IV, cause a 15,000- and 3,000-fold increase in DNA synthesis efficiency, respectively. However, both polymerases exhibit low processivity, adding 6 to 8 nucleotides before dissociating. Lesion bypass by pol V does not require β , γ -complex in the presence of non-hydrolysable ATP γ S, indicating that an intact RecA filament may be required for translesion synthesis.

DNA damage in *E. coli* triggers the induction of 'SOS' genes involved in DNA replication, repair and mutagenesis. These genes

are transcriptionally regulated by the LexA repressor protein, with RecA* acting as a co-protease mediating the self-cleavage and subsequent inactivation of LexA⁴. One facet of the SOS response is a large (~100-fold) increase in the rate of base substitution mutations targeted at template DNA damage sites, in addition to increased untargeted mutations at normal template sites. When pol V (UmuD'2C) is mutated, mutation rates in ultraviolet-irradiated cells drop to spontaneous background levels⁵. In addition to pol V, there are two other LexA-regulated *E. coli* DNA polymerases, pol II⁶ and pol IV³. Pol IV appears to have little, if any, role in causing damage-induced chromosomal mutations, but does cause mutations on lambda phage⁷ and F' episomes⁸. Enigmatic pol II⁹ has been shown to be crucial in error-free replication restart¹⁰. Our focus is on pols IV and V, which are error-prone in distinct ways.

Three common DNA lesions are thymine–thymine pyrimidine (6–4) pyrimidone photoproducts (TT (6–4) photoproducts) and thymine–thymine *cis-syn* cyclobutane photodimers (TT *cis-syn* photodimers) caused by ultraviolet radiation, and apurinic/apyrimidinic (abasic) sites arising from the spontaneous loss of a DNA base, or when glycosylases excise damaged bases or uracil from DNA¹¹. Replication *in vivo* is impeded by each lesion^{12–14}. We have compared the translesion synthesis (TLS) efficiency and base incorporation specificity of pol V, pol IV and pol III. The forms of the enzymes used were: pol V Mut (or mutasome¹⁵, consisting of pol V (UmuD'2C), RecA*, β , γ -complex and SSB); pol IV (DinB)³ with β , γ -complex and SSB; and pol III HE (holoenzyme of pol III core, β , γ -complex and SSB). Our measurements of TLS show that pol V Mut catalyses efficient bypass of all three lesions within 30 s (Fig. 1). The estimated lesion bypass rates are ~30, 20 and 80% min⁻¹ for the TT (6–4) photoproduct, TT photodimer and abasic moiety, respectively. No detectable bypass is observed in 30 s using pol III HE or pol IV; the latter shows incorporation opposite each of the lesions but extending no further. After 8 min faint bypass product bands appear for pol III HE and pol IV, with bypass rates of 1.2 and 0.7% min⁻¹, respectively, for the TT(6–4) photoproduct, 1 and 0.5% min⁻¹ for the TT photodimer, and 0.4 and 0.1% min⁻¹ for the abasic moiety. Extensions from a normal template position (T site, Fig. 1, left gel) are around 88, 98 and 68% at 30 s for pol V Mut, pol III HE and pol IV, respectively. The enzyme levels were chosen to allow similar primer utilization on undamaged DNA templates, with roughly 50% of each primer being extended at 16 min (Fig. 1, left gel). Observations at early times (<2 min) indicate that synthesis by pol III HE is processive, whereas that by pol V Mut or pol IV is much more distributive (Fig. 1).

We used a gel-fidelity assay¹⁶ to measure nucleotide incorporation specificities for pols III, IV and V opposite both positions of a TT

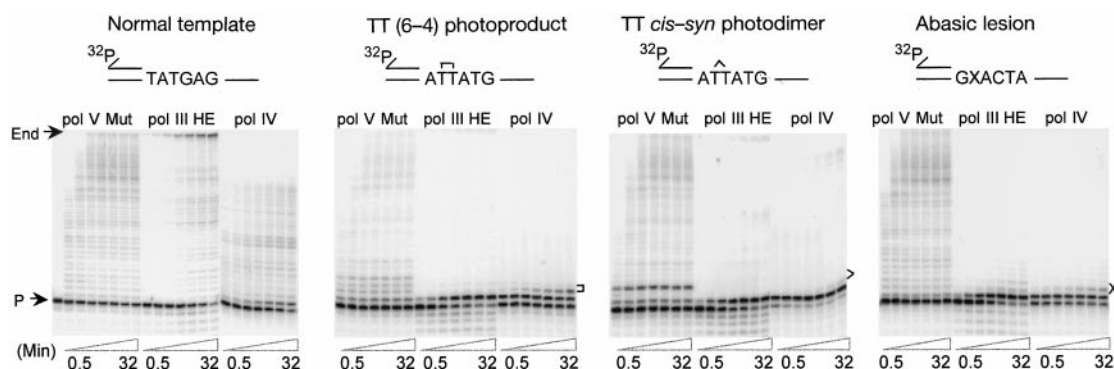


Figure 1 Comparison of translesion synthesis by pol III HE, pol IV and pol V Mut. Normal and translesion synthesis by pol III core (1 nM), pol IV (25 nM) and pol V (25 nM) were carried out in reactions containing 2 nM primer templates (shown at top), 1 mM ATP, 300 nM SSB, 40 nM β -subunit and 10 nM γ -complex. RecA protein (1 μ M) was present in pol V reactions to give mutasome, pol V Mut. Reactions were initiated by adding the

polymerase and four dNTPs (100 μ M each; see Methods). P indicates the location of the non-extended primer. The locations of a TT (6–4) photoproduct, TT *cis-syn* cyclobutane dimer and abasic site are indicated by the square bracket, diagonal bracket and cross, respectively. The lesion bypass rate is calculated as the fraction of primers extended beyond the lesion per min divided by the total amount of primer extended.

(6–4) photoproduct. A ^{32}P -labelled primer is extended by incorporating a single ‘running-start’ T to reach the 3′-T of a TT (6–4) photoproduct (Fig. 2a). Both pol V Mut and pol IV have no measurable 3′-exonuclease proofreading activity, and pol III α -HE is a proofreading-deficient form of pol III HE.

Mutations occur predominantly at the 3′-T site of a TT (6–4) photoproduct, with 5′-T mutations occurring much less frequently⁴. In accord with this observation, we find that pol V Mut favours incorporation of G over A opposite the 3′-T site by sixfold (Fig. 2a, Table 1). In contrast, pol III α -HE and pol IV favour incorporation of A over G by 27- and 3.3-fold, respectively (Fig. 2a). At the 5′-T, pol V Mut incorporates A about 8- to 13-fold more than G, when either A-T or G-T primer ends are extended (Fig. 2b, Table 1). Incorporation of G is not detectable at the 5′-T site with either pol III α -HE or pol IV (data not shown). Thus, our data for pol V Mut agree qualitatively and quantitatively with *in vivo* mutational data showing increased 3′-T $\rightarrow$ C transition mutations with a much smaller increase in 5′-T mutations^{14,17} (Table 1). The absence of detectable amounts of incorporation of either C or T at either TT site by pol V is consistent with the absence of transversions *in vivo* (data not shown). Similar kinetic analyses performed using a TT *cis-syn* photodimer and an abasic lesion also agree with *in vivo* measurements^{13,17,18} (Table 1). We conclude that pol V Mut is probably responsible for SOS mutations targeted to DNA damage

sites, on the basis of its high TLS efficiency and recapitulation of *in vivo* mutational specificity.

An examination of TLS gel bands also reveals the error-prone nature of pol V Mut at undamaged template sites (Fig. 2a). All three polymerases catalysed TLS by incorporating T opposite A and A opposite T at the two template sites immediately downstream from the lesion. However, pol V Mut also formed mispairs at the next template base G (Fig. 2a). These data are consistent with *in vivo* experiments showing that, in addition to mutations targeted specifically at template lesions, there are also untargeted pol V Mut mutations¹⁹. No such mispairs occurred with either pol IV or pol III α -HE.

We have also measured the fidelity of nucleotide incorporation of the three polymerases on undamaged DNA template sites (Table 2). For pol V Mut, we find error frequencies of $1\text{--}5 \times 10^{-3}$ for G-T, T-G, G-G, A-G and T-T mispairs; $3\text{--}7 \times 10^{-4}$ for G-A, C-A, C-T, T-C and A-C mispairs; 7×10^{-5} for A-A mispairs; and $<10^{-5}$ for C-C mispairs. For pol IV, the error frequencies ranged from $\sim 2 \times 10^{-3}$ for G-G to 3×10^{-5} for C-C mispairs. However, the G-G mispair is unlikely to be caused by direct misincorporation of dGMP opposite G, but rather by dNTP-stabilized misalignment²⁰ whereby incorporation of G occurs opposite a template C base immediately downstream from G. This is consistent with pol IV’s propensity for catalysing -1 frameshift errors³. For comparison with pols IV

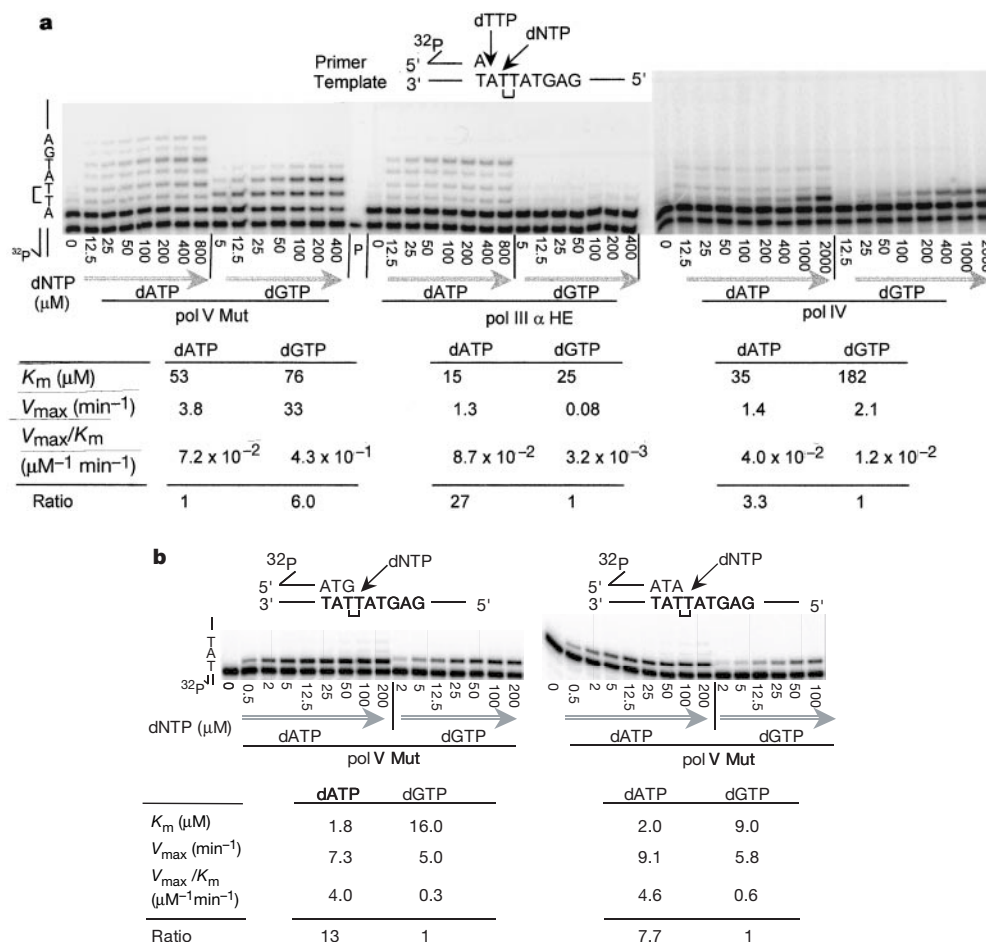


Figure 2 Incorporation of A compared with G opposite a TT (6–4) photoproduct. **a**, the incorporation kinetics of either A or G were measured opposite the 3′-T site of a TT (6–4) photoproduct using pol V Mut, pol IV or pol III α -HE. To reach the target site, a running-start T (10 μM dTTP) is incorporated opposite A as shown at top. The concentrations of dNTP = dATP or dGTP for incorporation opposite the target site were varied. P indicates the location of the non-extended primer. **b**, The incorporation kinetics of either A or G were

measured opposite the 5′-T of a TT (6–4) photoproduct using pol V Mut to extend a primer terminating with either G (left gel) or A (right gel) situated opposite the 3′-T of the TT (6–4) photoproduct. The kinetic analysis, summarized under Methods, is described in detail in ref. 16. Relative efficiencies of incorporation of dATP versus dGTP (ratios shown at the bottom of the tables) are fold-differences between $V_{\max}/K_m$ for incorporation of dATP compared to dGTP.

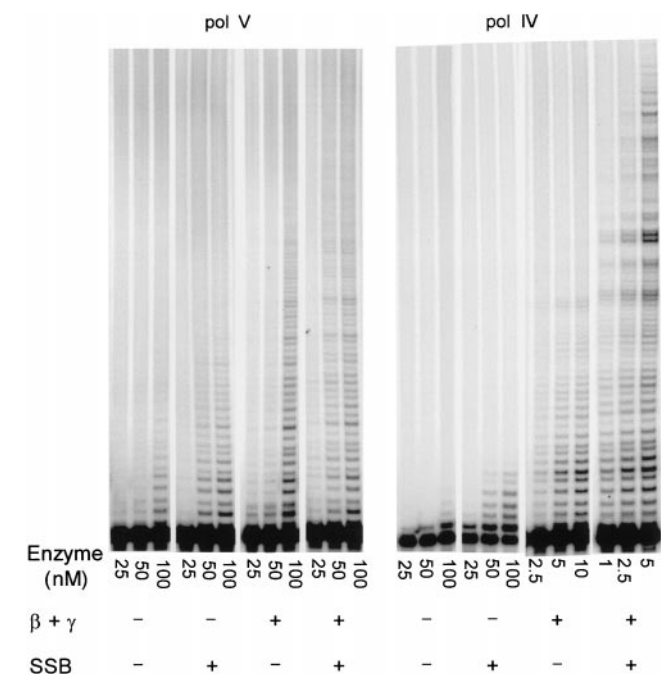


Figure 3 Analysis of the processivity of pol IV and pol V with or without SSB and β,γ -complex. The processivity of pol IV and pol V was measured on circular single-stranded DNA (M13 mp7) primed with a 5'- ^{32}P -labelled 30-mer (see Methods). Synthesis by pol V was carried out in the presence of RecA protein.

and V Mut, the proofreading-defective pol III α -HE makes base-substitution errors in the 10^{-4} to 10^{-6} range, consistent with previous measurements²⁰. As pols IV and V also exhibit an enhanced ability to extend mismatched primer ends^{1,3}, an increase in untar-geted mutations is likely to result from the combined effect of a relatively high error rate followed by efficient mismatch extension, which inhibits proofreading.

Synthesis on undamaged DNA templates by either pol V with RecA protein or pol IV alone results in a weak, distributive primer-elongation pattern, which is enhanced by addition of either β,γ -complex or SSB (Fig. 3). Synthesis by pol V in the absence of RecA protein, β,γ -complex and SSB is weak². Synthesis by pol IV is distributive in the absence of β,γ -complex³ (Fig. 3), and synthesis by pol V is distributive in the absence of RecA (data not shown). Maximal stimulation for both polymerases occurs when β,γ -complex and SSB are present simultaneously in the reaction (Fig. 3); but even under the most favourable conditions, processivities are limited to around 6–8 nucleotides for each polymerase. Similar results are obtained in the presence of a heparin trap (data not shown). Whereas RecA strongly stimulates pol V activity on both damaged and undamaged DNA templates and is required for pol V-catalysed TLS *in vitro*^{1,2}, it has no measurable effect on pol IV synthetic activity (data not shown).

Most striking, however, are the effects of RecA protein on pol V, and β,γ -complex on pol IV. There is a 15,000- and 3,000-fold increase in DNA synthesis efficiency (V_{max}/K_m) for pol V and pol IV, respectively (Fig. 4). These increased efficiencies are reflected in marked K_m reductions for dNTP substrates; the steady-state K_m values are reduced from 1.2 mM to 0.08 μM for pol V and from

Table 1 Pol V Mut insertion specificity at DNA template lesions: comparing *in vitro* and *in vivo* data

	TT (6–4) photoproduct						TT <i>cis-syn</i> photodimer						Abasic lesion	
	Opposite 3'T			Opposite 5'T			Opposite 3'T			Opposite 5'T				
	<i>In vitro</i> kinetics*	<i>In vivo</i> 1†	<i>In vivo</i> 2‡	<i>In vitro</i> kinetics	<i>In vivo</i> 1†	<i>In vivo</i> 2‡	<i>In vitro</i> kinetics	<i>In vivo</i> 1§	<i>In vivo</i> 2‡	<i>In vitro</i> kinetics	<i>In vivo</i> 1§	<i>In vivo</i> 2‡	<i>In vitro</i> kinetics	<i>In vivo</i>
A	14.3	11	27	92.2	94	98	98	94	98	95.2	99	98	63.6	54–80
G	85.7	87.5	69	7.8	2	2	2	1		4.8	<1	1	36.4	15–20

* The data represent the percentage of either A or G incorporated opposite each template lesion site derived from the ratio of A/G incorporation specificity (V_{max}/K_m) — see, for example, Fig. 2 for TT (6–4) photoproduct. Incorporation of either T or C opposite each of the lesions is negligible.

† From ref. 14.

‡ From ref. 17.

§ From ref. 18.

|| From ref. 13.

Table 2 Fidelity of pol III α HE, pol IV and pol V Mut by steady state kinetics

dNTP•Template	pol V Mut		pol IV		pol III α -HE	
	V_{max}/K_m ($\mu\text{M}^{-1}\text{min}^{-1}$)	f_{inc}^*	V_{max}/K_m ($\mu\text{M}^{-1}\text{min}^{-1}$)	f_{inc}^*	V_{max}/K_m ($\mu\text{M}^{-1}\text{min}^{-1}$)	f_{inc}^*
dGTP•G	0.23	2.7×10^{-3}	0.047	1.7×10^{-3}	0.00073	2.5×10^{-5}
dATP•G	0.11	1.3×10^{-3}	0.018	6.7×10^{-4}	0.0029	1.0×10^{-4}
dTTP•G	0.41	4.8×10^{-3}	0.023	8.5×10^{-4}	0.014	4.8×10^{-4}
dCTP•G	86	1	27	1	29	1
dGTP•A	0.033	3.3×10^{-4}	0.0031	1.5×10^{-4}	0.0038	3.2×10^{-5}
dATP•A	0.0070	7.0×10^{-5}	0.0011	5.2×10^{-5}	0.0011	9.2×10^{-6}
dTTP•A	100	1	21	1	120	1
dCTP•A	0.055	5.5×10^{-4}	0.0020	9.5×10^{-5}	0.019	1.6×10^{-4}
dGTP•T	0.91	2.4×10^{-3}	0.0027	3.6×10^{-4}	0.0035	2.5×10^{-5}
dATP•T	380	1	7.6	1	140	1
dTTP•T	1.4	3.7×10^{-3}	0.00067	8.8×10^{-5}	0.017	1.2×10^{-4}
dCTP•T	0.31	8.1×10^{-4}	0.0017	2.2×10^{-4}	0.0069	4.9×10^{-5}
dGTP•C	360	1	12	1	56	1
dATP•C	0.19	5.3×10^{-4}	0.0016	1.3×10^{-4}	0.0040	7.1×10^{-5}
dTTP•C	0.26	7.2×10^{-4}	0.0017	1.4×10^{-4}	0.00029	5.2×10^{-6}
dCTP•C	n.d.†	$< 1.0 \times 10^{-5}$	0.00043	3.6×10^{-5}	0.00016	2.9×10^{-6}

* The nucleotide misincorporation ratio, $f_{\text{inc}} = (V_{\text{max}}/K_m)_W / (V_{\text{max}}/K_m)_R$, where W and R denote incorporation of either a wrong or right nucleotide. The sensitivity of the assay is dependent on polymerase activity, $(V_{\text{max}}/K_m)_R$, and can detect a larger misincorporation range for pol III α -HE ($\sim 10^{-6}$) compared to pol V Mut ($< 10^{-6}$). The s.e. values for f_{inc} are $\pm 30\%$.

† n.d., not detected.

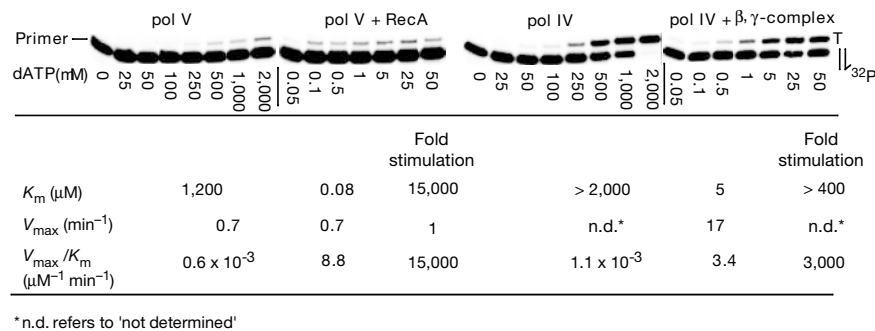


Figure 4 Effect of RecA and β, γ -complex on incorporation of correct dAMP at an undamaged template by pol V and pol IV. Reactions were carried out with undamaged circular ssDNA M13mp7 annealed to a 30-mer ^{32}P -labelled primer. Synthesis by pol V was performed in the presence of β, γ -complex, SSB and ATP. Apparent K_m and $V_{\max}$ values were obtained by plotting nucleotide incorporation rate against dNTP concentration

>2 mM to 5 μM for pol IV (Fig. 4). In agreement with a previous model²¹, it appears that RecA protein, and to a much lesser extent β, γ -complex, serves to target pol V to replication complexes stalled at DNA template lesions. The presence of RecA* is absolutely required for TLS by pol V^{1,2}. However, β, γ -complex can be dispensed with *in vitro* when non-hydrolysable ATP γS replaces ATP in the reaction (Fig. 5), indicating that RecA filament disassembly, occurring in the presence of ATP but not ATP γS ²², may 'extinguish' pol V activity. This result appears to resolve a difference in requirements for pol V-catalysed TLS between our data^{1,2} and those of ref. 23, which did not require β, γ -complex. In our experiments there is 1 RecA per 5 to 15 nucleotides, compared with a much higher 5 RecA per nucleotide in ref. 23, probably sufficient to maintain an intact RecA filament. Thus RecA may activate pol V while targeting it to a DNA damage site, whereas β, γ -complex could help tether pol V at the 3'-primer end during the ATP-driven dynamic RecA assembly-disassembly process. The large (3,000-fold) stimulation of β, γ -complex on pol IV activity indicates that processivity proteins might facilitate activation and targeting of pol IV at stalled replication forks on undamaged DNA.

Our *in vitro* data (Table 1) appear to recapitulate qualitatively and quantitatively the *in vivo* requirements of pol V Mut in causing SOS

and fitting the data to a saturation curve (rectangular hyperbola) using nonlinear least squares¹⁶. Note that for pol IV alone, individual K_m and $V_{\max}$ values could not be measured because the nucleotide incorporation rate remained linear throughout the entire dATP concentration range (0–2,000 μM); the $V_{\max}/K_m$ value is given by the slope of the line. n.d., not determined.

lesion-targeted mutations. The function of pol V Mut in generating SOS non-targeted chromosomal mutations may be to create genetic diversity during evolution²⁴. Proposing a biochemical role for pol IV during the SOS response seems more tenuous. It would be surprising if *E. coli* pol IV, an SOS-induced polymerase with many eukaryotic and prokaryotic DNA repair homologues^{25,26}, failed to have a well defined cellular function. Although one can rule out a role for pol IV in SOS-targeted ultraviolet mutagenesis⁴, it nevertheless causes untargeted mutations on phage lambda⁷, and also on an F' plasmid when overproduced⁸. Perhaps pol IV is required to alleviate stalled replication forks on undamaged DNA. Stalling of pol III HE can be caused by mismatched or misaligned primer-ends that cannot be proofread, a suggestion supported by the *in vitro* properties of pol III HE²⁰. Thus, error-prone pol IV and pol V may have complementary roles in *E. coli*, with pol IV extending aberrant primer 3'-ends and pol V Mut catalysing TLS. As pol IV and pol V Mut are both poorly processive and dissociate after relieving a blocked replication fork, pol III HE can then replace the errant polymerases to complete chromosomal replication. □

Methods

Materials

Ultrapure ATP, dNTPs, *E. coli* single-stranded binding protein (SSB) and RecA protein were purchased from Amersham-Pharmacia. Purification of pol III core and its accessory proteins²⁷ and pol V² were carried out as described. For DNA polymerase IV (DinB). The coding sequence of pol IV (DinB) was amplified from genomic *E. coli* DNA and cloned into pMAL-c2x (New England Biolabs) to create a Maltose-binding-protein–DinB overexpression vector. We purified the MBP–DinB fusion protein by amylose affinity chromatography using standard protocols. Fractions containing MBP–DinB were pooled, concentrated and applied to a gel-filtration Superdex 200 XK 26/60 column (Amersham-Pharmacia). The column was developed with buffer A (20 mM Tris-HCl (pH 7.4), 200 mM NaCl, 1 mM EDTA, 1 mM DTT). Fractions containing MBP–DinB (>95% pure) were collected, supplemented with 20% glycerol and aliquots frozen at -70°C .

DNA substrates

Templates used in this study were 7.2-kilobase (kb) linear M13mp7 single-stranded DNA with an abasic site 50 bases from the 5'-end¹⁵, and a TT (6–4) photoproduct or a TT *cis-syn* photodimer 53 bases from the 5' end. We prepared and characterized the TT lesions as described²⁸. 5'- ^{32}P -labelled 30-mers were used as primers. For the measurement of incorporation kinetics opposite an abasic site, the first T (3'-T) of a TT (6–4) photoproduct or a TT *cis-syn* photodimer, primers and templates were annealed, placing the 3'-primer end one base before the lesion site (running-start reaction¹⁶). To measure incorporation kinetics opposite the second T, the 3'-primer end is situated directly opposite the first T of the lesion (standing-start reaction¹⁶).

Nucleotide incorporation kinetics on lesioned and natural DNA primer templates

All reaction mixtures (10 μl) contained 20 mM Tris (pH 7.5), 8 mM MgCl_2 , 5 mM DTT, 0.1 mM EDTA, 25 mM sodium glutamate, 1 mM ATP, 40 $\mu\text{g ml}^{-1}$ BSA and 4% (v/v) glycerol. Primer/template DNA (2 nM) was incubated at 37°C for 2 min at varying dNTP substrate concentrations, 40 nM of β -subunit (as dimer), 10 nM γ -complex, 300 nM SSB (as tetramer) and RecA (1 μM) (for pol V, but not pol IV). Running-start reactions were initiated by adding 10 μM of a running start dNTP and either pol III α -subunit (1–5 nM),

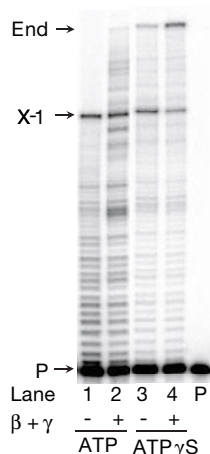


Figure 5 Role of β, γ -complex in abasic site bypass by pol V in the presence of ATP or ATP- γS . All lesion bypass reactions were carried out at 37°C for 10 min in the presence of 2 nM DNA substrate, RecA protein (1 μM), SSB (300 nM), pol V (100 nM), 4 dNTPs (100 μM) and 1 mM of either ATP or ATP- γS . β -subunit and γ -complex, if present, were at 40 nM and 10 nM, respectively. The DNA substrate is a ^{32}P -labelled primer annealed to a 7.2-kb linear single-stranded DNA with an abasic lesion, as described¹⁵. The unextended primer P, the upstream site before the lesion (X-1) and the end of template are indicated on the left.

pol IV (20 nM) or pol V (25–50 nM). The running-start nucleotide was absent in standing-start reactions. Reactions were run at 37 °C for 5 min and quenched by adding 20 µl of 20 mM EDTA in 95% formamide. Primer–template extension products were heat denatured and run on a 16% polyacrylamide denaturing gel. We measured integrated gel-band intensities by phosphorimaging using ImageQuant software (Molecular Dynamics). Nucleotide incorporation rates opposite both undamaged and lesion target sites were determined, and apparent V_{max}/K_m values and nucleotide misincorporation frequencies were calculated as described¹⁶. We carried out fidelity measurements using standing-start reactions on undamaged linear M13mp7 DNA templates annealed to 30mer ³²P-labelled primers with ATP (100 µM). The following template sequences were used: 3'...GTTGC CG...; 3'...CGTAGCC...; 3'TCGTAGCC...; 3'TATCAAC, where the base in bold is the template target site at which the nucleotide misincorporation ratio, f_{inc} , was calculated.

Processivity measurements

We measured the processivity of pol IV and pol V Mut on a single-stranded circular DNA (M13 mp7) primed with a 5'-³²P-labelled 30-mer primer as described²⁰. A fixed amount of substrate (10 nM) was incubated with decreasing amounts of pol IV or pol V at 37 °C for 3 min. If indicated, other proteins included in the reactions were: RecA (5 µM), SSB (1.5 µM), β (200 nM) and γ-complex (50 nM). ATP (1 mM) is added in reactions containing RecA and/or γ-complex. Reactions were initiated by adding four dNTPs (100 µM), quenched after 5 min and analysed on a 12% denaturing polyacrylamide gel. Reaction products were quantified by phosphorimaging. Processivities were measured at polymerase concentrations, where product lengths remained stable, and the number of primers used remained less than 10%. We computed processivity as a ratio of the total number of nucleotides added in the reaction to the number of extended DNA primers. High enzyme/DNA ratios are required to carry out primer extension (Fig. 3), probably because the polymerases bind to the long regions of single-stranded DNA. We have previously observed cooperative binding of pol V to single-stranded DNA²⁰.

Received 13 December 1999; accepted 4 February 2000.

1. Tang, M. *et al.* Biochemical basis of SOS mutagenesis in *Escherichia coli*: reconstitution of *in vitro* lesion bypass dependent on the UmuD₂C mutagenic complex and RecA protein. *Proc. Natl Acad. Sci. USA* **95**, 9755–9760 (1998).
2. Tang, M. *et al.* UmuD₂C is an error-prone DNA polymerase, *Escherichia coli* pol V. *Proc. Natl Acad. Sci. USA* **95**, 8919–8924 (1999).
3. Wagner, J. *et al.* The *dinB* gene encodes a novel *Escherichia coli* DNA polymerase (DNA pol IV). *Mol. Cell* **40**, 281–286 (1999).
4. Friedberg, E. C., Walker, G. C. & Siede, W. *DNA Repair and Mutagenesis* 407–522 (ASM, Washington, 1995).
5. Kato, T. & Shinoura, Y. Isolation and characterization of mutants of *Escherichia coli* deficient in induction of mutagenesis by ultraviolet light. *Mol. Gen. Genet.* **156**, 121–131 (1977).
6. Bonner, C. A., Hays, S., McEntee, K. & Goodman, M. F. DNA polymerase II is encoded by the DNA damage-inducible *dinA* gene of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **87**, 7663–7667 (1990).
7. Brotcorne-Lannoye, A. & Maenhaut-Michel, G. Role of RecA protein in untargeted UV mutagenesis of bacteriophage λ: Evidence for the requirement for the *dinB* gene. *Proc. Natl Acad. Sci. USA* **83**, 3904–3908 (1986).
8. Kim, S.-R. *et al.* Multiple pathways for SOS-induced mutagenesis in *Escherichia coli*: An over-expression of *dinB/dinP* results in strongly enhancing mutagenesis in the absence of any exogenous treatment to damage DNA. *Proc. Natl Acad. Sci. USA* **94**, 13792–13797 (1997).
9. Knippers, R. DNA polymerase II. *Nature* **228**, 1050–1053 (1970).
10. Rangarajan, S., Woodgate, R. & Goodman, M. F. A phenotype for enigmatic DNA polymerase II: a pivotal role for pol II in replication restart in UV-irradiated *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **96**, 9224–9229 (1999).
11. Lindahl, T. DNA repair enzymes. *Annu. Rev. Biochem.* **51**, 61–87 (1982).
12. Banerjee, S. K., Christensen, R. B., Lawrence, C. W. & LeClerc, J. E. Frequency and spectrum of mutations produced by a single *cis-syn* thymine–thymine cyclobutane dimer in a single-stranded vector. *Proc. Natl Acad. Sci. USA* **85**, 8141–8145 (1988).
13. Lawrence, C. W., Borden, A., Banerjee, S. K. & LeClerc, J. E. Mutation frequency and spectrum resulting from a single abasic site in a single-stranded vector. *Nucleic Acids Res.* **18**, 2153–2157 (1990).
14. LeClerc, J. E., Borden, A. & Lawrence, C. W. The thymine–thymine pyrimidine–pyrimidone (6–4) ultravioletlight photoproduct is highly mutagenic and specifically induces 3' thymine-to-cytosine transitions in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **88**, 9685–9689 (1991).
15. Rajagopalan, M. *et al.* Activity of the purified mutagenesis proteins UmuC, UmuD', and RecA in replicative bypass of an abasic DNA lesions by DNA polymerase III. *Proc. Natl Acad. Sci. USA* **89**, 10777–10781 (1992).
16. Creighton, S., Bloom, L. B. & Goodman, M. F. in *Methods Enzymol.* Vol. 262 (ed. Campbell, J. L.) 232–256 (Academic, San Diego, 1995).
17. Smith, C. A. *et al.* Mutation spectra of M13 vectors containing site-specific *cis-syn*, *trans-syn-I*, (6–4), and dewar pyrimidine photoproducts of thymidyl-(3'→5')-thymidine in *Escherichia coli* under SOS conditions. *Biochemistry* **35**, 4146–4154 (1996).
18. Lawrence, C. W., Banerjee, S. K., Borden, A. & LeClerc, J. E. T–T Cyclobutane dimers are misinstructive, rather than non-instructive, mutagenic lesions. *Mol. Gen. Genet.* **166**, 166–168 (1990).
19. Fijalkowska, I. J., Dunn, R. L. & Schaaper, R. M. Genetic requirements and mutational specificity of the *Escherichia coli* SOS mutator activity. *J. Bacteriol.* **179**, 7435–7445 (1997).
20. Bloom, L. B. *et al.* Fidelity of *Escherichia coli* DNA polymerase III holoenzyme: The effects of β, γ complex processivity proteins and ε proofreading exonuclease on nucleotide misincorporation efficiencies. *J. Biol. Chem.* **272**, 27919–27930 (1997).
21. Sommer, S., Boudsocq, F., Devoret, R. & Bailone, A. Specific RecA amino acid changes affect RecA–UmuD' interaction. *Mol. Microbiol.* **28**, 281–291 (1998).
22. Menetski, J. P. & Kowalczykowski, S. C. Interaction of RecA protein with single-stranded DNA: Quantitative aspects of binding affinity modulation by nucleotide cofactors. *J. Mol. Biol.* **181**, 281–295 (1985).
23. Reuven, N. B., Arad, G., Maor–Shoshani, A. & Livneh, Z. The mutagenic protein UmuC is a DNA

polymerase activated by UmuD', RecA, and SSB and is specialized for translesion replication. *J. Biol. Chem.* **274**, 31763–31766 (1999).

24. Radman, M. Enzymes of evolutionary change. *Nature* **401**, 866–869 (1999).
25. McDonald, J. P. *et al.* Novel human and mouse homologs of *Saccharomyces cerevisiae* DNA polymerase η. *Genomics* **60**, 20–30 (1999).
26. Gerlach, V. L. *et al.* Human and mouse homologs of *Escherichia coli* DinB (DNA polymerase IV), members of the UmuC/DinB superfamily. *Proc. Natl Acad. Sci. USA* **96**, 11922–11927 (1999).
27. Naktinis, V., Turner, J. & O'Donnell, M. A molecular switch in a replication machine defined by an internal competition for protein rings. *Cell* **84**, 137–145 (1996).
28. Jing, Y., Kao, J. F.-L. & Taylor, J.-S. Thermodynamic and base-pairing studies of matched and mismatched DNA dodecamer duplexes containing *cis-syn*, (6–4) and Dewar photoproducts of TT. *Nucleic Acids Res.* **26**, 3845–3853 (1998).
29. Bambara, R. A., Fay, P. J. & Mallaber, L. M. Methods of analyzing processivity. *Methods Enzymol.* **262**, 270–280 (1995).
30. Bruck, L., Woodgate, R., McEntee, K. & Goodman, M. F. Purification of a soluble UmuD' C complex from *Escherichia coli*. Cooperative binding of UmuD' C to single-stranded DNA. *J. Biol. Chem.* **271**, 10767–10774 (1996).

Acknowledgements

This work was supported by National Institutes of Health grants to M.F.G., M.O. and J.-S.T.).

Correspondence and requests for materials should be addressed to M.F.G. (e-mail: mgoodman@mizar.usc.edu).

Retrotransposition of a bacterial group II intron

Benoit Cousineau, Stacey Lawrence, Dorie Smith & Marlene Belfort

Molecular Genetics Program, Wadsworth Center, New York State Department of Health, and School of Public Health, State University of New York at Albany, PO Box 22002, Albany, New York 12201-2002, USA

Self-splicing group II introns may be the evolutionary progenitors of eukaryotic spliceosomal introns^{1–7}, but the route by which they invade new chromosomal sites is unknown. To address the mechanism by which group II introns are disseminated, we have studied the bacterial Ll.LtrB intron from *Lactococcus lactis*⁸. The protein product of this intron, LtrA, possesses maturase, reverse transcriptase and endonuclease enzymatic activities^{9–11}. Together with the intron, LtrA forms a ribonucleoprotein (RNP) complex which mediates a process known as retrohoming¹¹. In retrohoming, the intron reverse splices into a cognate intronless DNA site. Integration of a DNA copy of the intron is recombinase independent but requires all three activities of LtrA¹¹. Here we report the first experimental demonstration of a group II intron invading ectopic chromosomal sites, which occurs by a distinct retrotransposition mechanism. This retrotransposition process is endonuclease-independent and recombinase-dependent, and is likely to involve reverse splicing of the intron RNA into cellular RNA targets. These retrotranspositions suggest a mechanism by which spliceosomal introns may have become widely dispersed.

The group II intron donor in these experiments is a twintron/Kan^R variant of Ll.LtrB, carrying the self-splicing group I *td* intron and a kanamycin resistance (Kan^R) gene (Fig. 1a, b)¹¹. The Kan^R marker facilitates selection of group II intron mobility events. The group I intron allows identification of those mobility events that have proceeded through an RNA intermediate and in which the group I intron has therefore been lost. The donor was first tested for retrohoming using *L. lactis* and a two-plasmid system with a recipient plasmid containing the cognate Ll.LtrB homing site (Fig. 1a)¹¹. After RNP expression under the control of a nisin-inducible promoter¹², inheritance of the Kan^R-marked intron occurred at 40% or 58% per recipient in two different wild-type hosts (Fig. 1c, crosses 1 and 3, H%). Of these homing products, 98%

had lost the group I intron (Fig. 1c, crosses 1 and 3, GI loss %), indicating that retrohoming was at least 39% or 57% in the two *L. lactis* hosts (Fig. 1c, crosses 1 and 3, RH%). Thus, retrohoming of Ll.LtrB in *L. lactis* is extremely efficient, with a tenfold increase over uninduced levels¹¹ of the RNP, and a greater than 100-fold increase in Kan^R recombinants above a donor expressing non-functional LtrA (Fig. 1c, compare crosses 3 and 4). Retrohoming levels in an Endo⁻ mutant that retains reverse transcriptase and maturase activity

(J. San Filippo and A. M. Lambowitz, unpublished) dropped to 2% of the LtrA⁺ parent (Fig. 1c, compare crosses 3 and 5). This agrees with the behaviour of this mutant in *Escherichia coli* (5% of LtrA⁺, J. San Filippo and A. M. Lambowitz, unpublished). Furthermore, retrohoming levels remained high in $\Delta recA$ cells lacking an intact homologous recombination system, at 30% per recipient, or 74% of the efficiency in the *recA*⁺ wild-type host (Fig. 1c, compare crosses 1 and 2). These results are consistent with an endonuclease-mediated recombination-independent mechanism¹¹.

The donor plasmid for detecting group II intron movement to ectopic chromosomal sites has the nisin-inducible twintron/Kan^R expression cassette in an erythromycin-resistant (Erm^R) temperature-sensitive (Ts) shuttle vector. This plasmid is maintained in *L. lactis* at 30 °C but cannot replicate at 37 °C (Fig. 1b¹³). Therefore Kan^R transposition events into the chromosome were selected after increasing the temperature and then screened for plasmid curing (Erm^S) and the presence of both introns. Genomic DNA containing retrotransposed group II introns was then isolated from colonies that were Kan^R-Erm^S, group II intron positive and group I intron negative. After cloning the Kan^R genomic fragments, we determined the insertion sites of the intron by sequencing the flanking DNA.

We characterized eight different ectopic insertion sites of Ll.LtrB in the chromosome of *L. lactis* (Fig. 2, 1–8). Two of these sites were at different positions in 23S ribosomal RNA. Also, one was in the *hdsM* gene (methylase; Genbank AF034786), four were in open reading frames (ORFs) of unknown function and one was in undefined sequence (D. Ehrlich and A. Sorokin, personal communication). Remarkably, insertions 1–7 were all in the sense orientation of the gene.

Analysis of the sequences flanking retrotransposed introns indicates that the base pairings between the EBS1 sequence of the intron

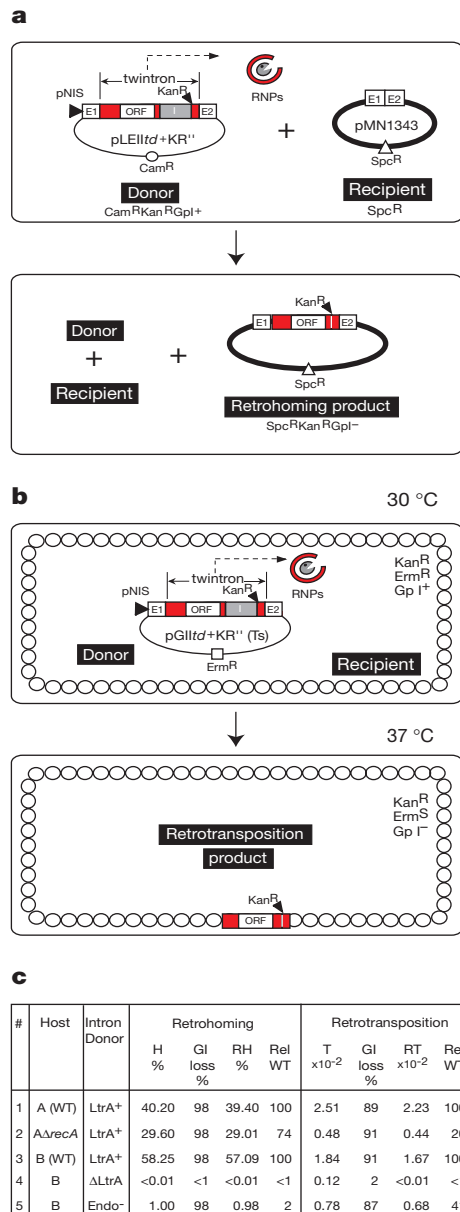


Figure 1 Retrohoming and retrotransposition in *L. lactis*. **a**, Retrohoming assay. Ll.LtrB group II intron, red; group I *td* intron, grey; E1 and E2, Ll.LtrB exons; RNPs, LtrA protein and intron RNA; pNIS, nisin-inducible promoter. **b**, Retrotransposition assay. Recipient is bacterial chromosome. **c**, Retrohoming and retrotransposition levels. H, homing, $H = \text{Kan}^R\text{Cam}^R/\text{Cam}^R \times 100$ (ref. 11); GI loss, group I intron loss; RH, retrohoming, $RH = H \times \text{GI loss}$; T, transposition, RT, retrotransposition, $RT = T \times \text{GI loss}$; Rel WT = RH or RT relative to 100 for wild-type host (A or B) and intron donor (LtrA⁺). T and RT are expressed per total cell number (although relative values in different crosses are accurate, absolute values may be elevated by Kan^R enrichment). All assays were repeated three times concomitantly. Standard errors for H% are: cross 1, 5.92; cross 2, 3.60; cross 3, 5.22; cross 4, not calculated; cross 5, 0.13. Standard errors for $T \times 10^{-2}$ are: cross 1, 0.62; cross 2, 0.02; cross 3, 1.06; cross 4, 0.03; cross 5, 0.32.

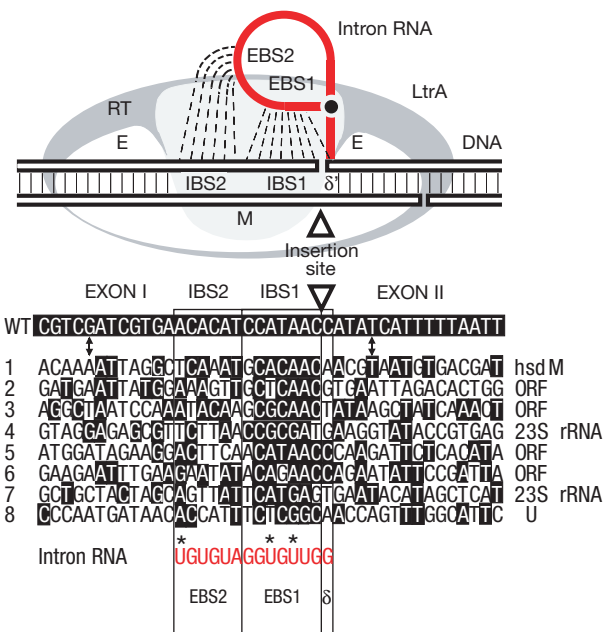


Figure 2 Ectopic insertions of Ll.LtrB in the *L. lactis* chromosome. Top, Interaction between group II intron RNP and target DNA. Dotted lines, base pairings (EBS1/IBS1 and EBS2/IBS2) between intron RNA (red lariat) and DNA target; breaks in each DNA strand, cleavage sites. Three activities of LtrA protein: reverse transcriptase, RT; endonuclease, E; maturase, M. Bottom, wild-type retrohoming site (WT) and insertion sites of eight ectopic integrants (1–8). White nucleotides on black background represent identity with wild type, or the possibility of intron RNA interacting via G–U base pairs (asterisk). U, undefined sequence. Bidirectional arrows point at protein recognition sites G–21 and T+5.

and the IBS1 sequence 5' to the intron insertion site determine the specificity of LL.LtrB integration (Fig. 2). Conservation of nucleotides indicates that LL.LtrB may invade new sites by reverse splicing downstream of locations resembling its natural exon I sequence, and that the IBS1/EBS1 interaction is more important than the IBS2/EBS2 interaction, as in fungal systems^{14,15}. In accord with reverse splicing into the ectopic target sites, most of the retrotransposed introns were also able to undergo some degree of forward splicing, as determined by primer extension analysis of *L. lactis* RNA containing retrotransposed LL.LtrB. With an intron-specific primer (Fig. 3)¹⁰, complementary DNA corresponding to excised intron lariat, an indicator of splicing, was observed in most cases. Where the lariat is not detected there may be an absence of splicing, or simply low expression of the interrupted gene. The ectopic insertion of LL.LtrB near the 5' end of a 23S rRNA gene (Fig. 2, site 7) allowed us to estimate the splicing efficiency of LL.LtrB from this site, because the unspliced precursor was detectable. The ratio of spliced intron lariat to total RNA (precursor + lariat) was estimated at 0.1, compared with 0.8 for the intron at its native site.

Significantly, determinants of LtrA protein binding in the exons are not discernable. In retrohoming, protein recognition of the DNA target site is required at position -21(G) for reverse splicing and at position +5(T) for antisense-strand cleavage (Fig. 2, bottom)¹⁶. The lack of apparent protein recognition sites, particularly at T+5 in exon II, which is required for LtrA-mediated cleavage, suggests that retrotransposition might occur by an alternative pathway to retrohoming, a pathway where the LtrA endonuclease activity is dispensable.

To explore different pathways in ectopic intron mobility, we performed retrotransposition assays in the presence and absence of *L. lactis* RecA function and LtrA endonuclease activity. Whereas retrohoming proceeded at 74% of wild-type levels in *recA*⁻ cells, retrotransposition of LL.LtrB in the *recA*⁻ strain was reduced to 20% of that in a *recA*⁺ host (Fig. 1c, compare crosses 1 and 2, Rel WT), indicating that, unlike retrohoming, a significant proportion of retrotransposition events followed a recombination-dependent pathway. Furthermore, the LtrA endonuclease mutant, which strongly inhibited retrohoming (2% of the LtrA⁺ donor), was

greater than 20-fold more active in retrotransposition (41% of the LtrA⁺ donor; Fig. 1c, compare crosses 3 and 5).

Therefore only a fraction (20%) of retrotransposition events appear to occur via an illegitimate RecA-independent retrohoming mechanism (Fig. 4a). The retrohoming-like pathway, involving reverse splicing into double-stranded DNA (step 1a), cDNA synthesis (steps 2a and 3a), second-strand synthesis (step 4a) and ligation (step 5a), would be independent of homologous recombination, but dependent on the endonuclease activity of LtrA (step 1a)¹¹. This minor pathway would explain the endonuclease requirement of some retrotransposition events (Fig. 1c, compare crosses 3 and 5). However, the endonuclease-independence of 41% of retrotransposition events indicates that second-strand cleavage may not be required and that reverse splicing might occur into single-stranded nucleic acid, either single-stranded DNA (for example at a replication fork or an actively transcribed gene) or RNA. Whereas integration into DNA would probably involve the replication machinery but not recombination, integration into RNA is predicted to be RecA-dependent (Fig. 4b), as observed for most retrotransposition events. Reverse-splicing into RNA (step 1b) would require recombinase function, as for example after cDNA synthesis (steps 2b and 3b), where the double-stranded DNA undergoes a double crossover event (step 4b). Resolution (step 5b) would yield the retrotransposition product, with the intronless chromosomal allele being replaced by the intron-containing allele.

Interestingly, residual retrotransposition events in a *recA*⁻ background (20%; Fig. 1c, cross 2) that typify the DNA-based pathway

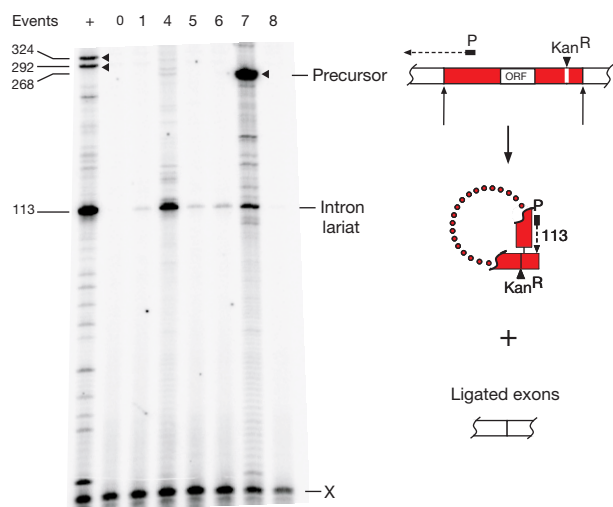


Figure 3 Splicing of LL.LtrB at ectopic sites. Primer extension was performed on total RNA¹⁰ from strain NZ9800ΔLL.LtrB containing ectopic chromosomal insertions (grown at 25 °C). Lengths of cDNA bands, synthesized with intron-specific primer (P), are indicated in nucleotides on the left. They correspond to precursor, excised intron lariat and an unknown *L. lactis* product (X) used as loading control. The positive control (+) has two independent transcription starts (black arrowheads)^{10,11}. 0, intronless *L. lactis* (NZ9800ΔLL.LtrB); lane numbers correspond to insertion sites in Fig. 2. Dashed lines, cDNAs; black bar, primer; upward-directed arrows, splice sites.

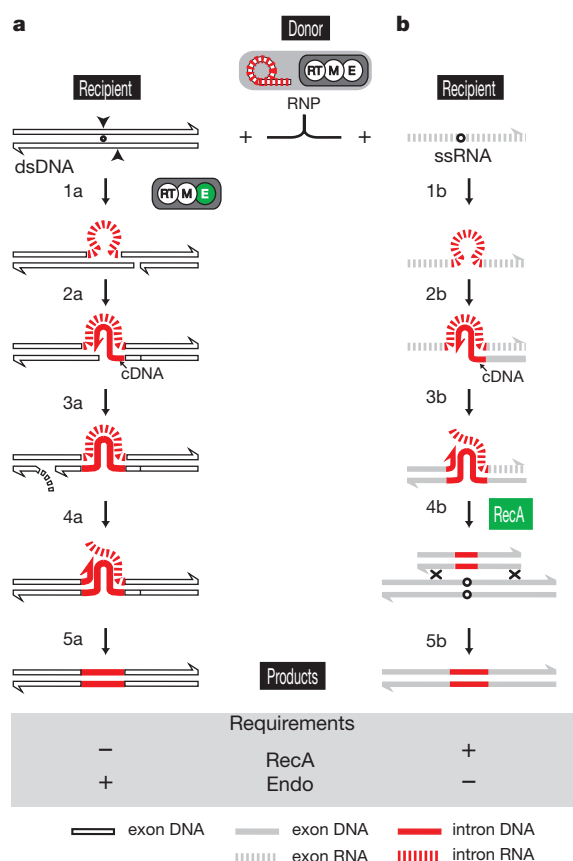


Figure 4 Mobility pathways. RNPs interact and promote reverse splicing of the intron into (a) dsDNA and (b) RNA. The RNP is depicted on a grey background with the intron lariat (red) and the trifunctional protein, LtrA, with reverse transcriptase (RT), endonuclease (E) and maturase (M) activities. Functions required exclusively by one pathway are shown in green. a, 1a, reverse splicing into dsDNA and endonuclease cleavage; 2a and 3a, cDNA synthesis; 4a, second-strand synthesis; 5a, repair. b, 1b, reverse splicing into RNA; 2b, cDNA synthesis; 3b, second-strand synthesis; 4b, recombination; 5b, resolution.

(Fig. 4a), plus those events with an Endo⁻ mutant (41%; Fig. 1c, cross 5) that are characteristic of the RNA-based pathway (Fig. 4b), fall short of 100%. This observation may be due to compromised reverse transcriptase and maturase activity in the endonuclease mutant, or to additional pathways to those shown in Fig. 4.

Retrotransposition into an RNA target is not only supported by the RecA-dependence and endonuclease-independence of many of the events, but is also consistent with the following observations. First, all the ectopic insertions are in the sense orientation of the respective genes, in accord with reverse splicing into expressed RNA (Fig. 2). Second, two of eight events are into a 23S rRNA gene. Ribosomal RNAs in a variety of organisms from all three kingdoms contain many introns at different positions, so these abundant RNAs provide many targets for reverse splicing and intron dissemination¹⁷. Third, autocatalytic introns are able to reverse-splice into RNA *in vitro* and *in vivo*^{17–23}. Regardless of pathway, the dispensability of the endonuclease activity of LtrA in retrotransposition reduces the specificity of site recognition by the RNPs to a small number of base pairs, as few as 8 out of 13 of the IBS1-IBS2 residues (Fig. 2), thereby promoting the promiscuity of group II introns and their ability to invade new sites.

The mechanism of DNA-targeted retrohoming and target-primed cDNA synthesis of group II introns (Fig. 4a) has led to comparisons to non-long-terminal-repeat (LTR) retrotransposons, which also encode proteins with reverse transcriptase and endonuclease activities^{24,25}. However, there are clear differences between retrotransposition of group II introns and non-LTR retrotransposons. First, there are striking mechanistic distinctions, with RNA being a possible target of a retrotransposing group II intron, and DNA being the exclusive target of a non-LTR retrotransposon. Second, reverse splicing of group II introns during retrotransposition guarantees a basic level of splicing at the newly invaded site (Fig. 3). Such events would be less deleterious to the host than gene invasion by non-LTR retrotransposons, thereby facilitating group II intron spread.

The discovery of group II introns in bacteria reinforced previous arguments for the antiquity of these catalytic RNA elements^{3,26}. The group II introns might have been pivotal in the evolution of eukaryotic genomes. On the one hand, the parallels in target-primed cDNA-synthesis and associated protein enzymes indicate that they may share evolutionary history with non-LTR retrotransposons^{24,25,27,28}, which occupy >17% of the human genome²⁹. On the other hand, because of similarities in splicing pathways and RNA components, group II introns may be ancestral to the spliceosome-dependent nuclear introns², which form >16% of the human genome (N. J. Schisler and J. Palmer, personal communication). Because of likeness in splicing chemistry, the mode of retrotransposition of group II introns in bacteria by reverse splicing into RNA may provide clues to the means of dispersal of spliceosomal introns. As an RNA-targeted mechanism is independent of endonuclease function, and because reverse splicing into RNA ensures forward splicing to maintain gene expression, this pathway is the most plausible mode of dissemination of spliceosomal introns in modern genomes of extant eukaryotes. □

Methods

Retrohoming assay

Spc^RKan^R products were selected from crosses between Cam^RKan^R donors (pLEIt^dKR⁺) and Spc^R recipients (pMN1343)¹¹. Cells were grown at 30 °C until the optical density at 600 nm was 0.5, then induced with nisin¹² for 3 h before plating under selective conditions. Assays were performed in *L. lactis* strain NZ9800 (wild-type host A) or NZ9800ΔLLtrB lacking the resident LLtrB intron (wild-type host B). Intron donors were wild-type (LtrA⁺), carried an LtrA deletion that removes all three activities (ΔLtrA, a >500-amino-acid deletion designated ΔORF¹⁰) or an LtrA mutant lacking endonuclease activity (Endo⁻ YRT mutant of J. San Filippo and A. M. Lambowitz, personal communication). Spc, spectinomycin; Cam, chloramphenicol.

Retrotransposition assay

The donor plasmid pGIIt^dKR⁺, a twintron/Kan^R derivative of temperature-sensitive pG⁺host5 (Erm^R)¹³, was transformed into *L. lactis*. Transformants were grown at 30 °C, and RNPs induced with nisin. After 3 h, cultures were diluted 100-fold into fresh medium and shifted to 37 °C for about 24 h to inhibit plasmid replication and stimulate plasmid loss. Retrotransposition events were enriched after 100-fold dilution in GM17-Kan medium (Kan^R) by overnight growth at 37 °C, and selected on kanamycin-containing plates. Kan^RErm^R colonies were analysed for group II intron acquisition and group I intron loss with ³²P-labelled region-specific dsDNA probes.

Received 2 December 1999; accepted 10 February 2000.

- Jacquier, A. Self-splicing group II and nuclear pre-mRNA introns: how similar are they? *Trends Biochem. Sci.* **15**, 351–354 (1990).
- Sharp, P. A. Five easy pieces. *Science* **254**, 663 (1991).
- Cavalier-Smith, T. Intron phylogeny: a new hypothesis. *Trends Genet.* **7**, 145–148 (1991).
- Roger, A. J. & Doolittle, W. F. Why introns-in-pieces? *Nature* **364**, 289–290 (1993).
- Weiner, A. M. mRNA splicing and autocatalytic introns: distant cousins or the products of chemical determinism? *Cell* **72**, 161–164 (1993).
- Michel, F. & Ferat, J. Structure and activities of group II introns. *Annu. Rev. Biochem.* **64**, 435–461 (1995).
- Hetzer, M., Wurzer, G., Schwyen, R. J. & Mueller, M. W. Trans-activation of group II intron splicing by nuclear U5 snRNA. *Nature* **386**, 417–420 (1997).
- Mills, D. A., Manias, D. A., McKay, L. L. & Dunny, G. M. Homing of a group II intron from *Lactococcus lactis* subsp. *lactis* ML3. *J. Bacteriol.* **179**, 6107–6111 (1997).
- Lambowitz, A. M., Caprara, M. G., Zimmerly, S. & Perlman, P. S. In *The RNA World* 451–485 (Cold Spring Harbor Laboratory Press, New York, 1999).
- Matsuura, M. *et al.* A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. *Genes Dev.* **11**, 2910–2924 (1997).
- Cousineau, B. *et al.* Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell* **94**, 451–462 (1998).
- Kuipers, O. P., Beerthuyzen, M. M., de Ruyter, P. G., Luesink, E. J. & de Vos, W. M. Autoregulation of nisin biosynthesis in *Lactococcus lactis* by signal transduction. *J. Biol. Chem.* **270**, 27299–27304 (1995).
- Biswas, I., Gruss, A., Ehrlich, S. D. & Maguin, E. High-efficiency gene inactivation and replacement system for gram-positive bacteria. *J. Bacteriol.* **175**, 3628–3625 (1993).
- Mueller, M. W., Allmaier, M., Eskes, R. & Schwyen, R. J. Transposition of group II intron *all* in yeast and invasion of mitochondrial genes at new locations. *Nature* **366**, 174–176 (1993).
- Sellem, C. H., Lecellier, G. & Belcour, L. Transposition of a group II intron. *Nature* **366**, 176–178 (1993).
- Mohr, G., Smith, D., Belfort, M. & Lambowitz, A. M. Rules for DNA target site recognition by a *Lactococcus* intron enable retargeting of the intron to specific DNA sequences. *Genes Dev.* **14**, 559–573 (2000).
- Roman, J. & Woodson, S. A. Reverse splicing of the *Tetrahymena* IVS: Evidence for multiple reaction sites in the 23S tRNA. *RNA* **1**, 478–490 (1995).
- Woodson, S. A. & Cech, T. R. Reverse self-splicing of the *Tetrahymena* group I intron: Implication for the directionality of splicing and for intron transposition. *Cell* **57**, 335–345 (1989).
- Augustin, S., Muller, M. W. & Schwyen, R. J. Reverse self-splicing of group II intron RNAs *in vitro*. *Nature* **343**, 383–386 (1990).
- Mori, M. & Schmelzer, C. Integration of group II intron *bII* into a foreign RNA by reversal of the self-splicing reaction *in vitro*. *Cell* **60**, 629–636 (1990).
- Mohr, G. & Lambowitz, A. M. Integration of a group I intron into a ribosomal RNA sequence promoted by a tyrosyl-tRNA synthetase. *Nature* **354**, 164–167 (1991).
- Thompson, A. J. & Herrin, D. L. A chloroplast group I intron undergoes the first step of reverse splicing into host cytoplasmic 5.8S rRNA. *J. Mol. Biol.* **236**, 455–468 (1994).
- Roman, J., Rubin, M. N. & Woodson, S. A. Sequence specificity of *in vivo* reverse splicing of the *Tetrahymena* group I intron. *RNA* **5**, 1–13 (1999).
- Eickbush, T. H. Mobile introns: retrohoming by complete reverse splicing. *Curr. Biol.* **9**, R11–R14 (1999).
- Curcio, M. J. & Belfort, M. Retrohoming: cDNA-mediated mobility of group II introns requires a catalytic RNA. *Cell* **84**, 9–12 (1996).
- Ferat, J. & Michel, F. Group II self-splicing introns in bacteria. *Nature* **364**, 358–361 (1993).
- Xiong, Y. & Eickbush, T. H. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* **9**, 3353–3362 (1990).
- Zimmerly, S., Guo, H., Perlman, P. S. & Lambowitz, A. M. Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* **82**, 545–554 (1995).
- Smit, A. F. The origin of interspersed repeats in the human genome. *Curr. Opin. Genet. Dev.* **6**, 743–748 (1996).

Acknowledgements

We thank J. San Filippo and A. Lambowitz for providing the LtrA Endo⁻ YRT mutant. We also thank D. Manias and G. Dunny for help in constructing pLEIt^dKR⁺; D. Ehrlich and A. Sorokin for information based on the *L. lactis* DNA sequence; J. Curcio, K. Derbyshire, V. Derbyshire, S. Hanes, A. Lambowitz, R. Lease, R. Morse, D. Nag and M. Parker for comments on the manuscript; N. J. Schisler and J. Palmer for estimates of intron composition of the human genome; and M. Carl for manuscript preparation. This work was supported by NIH grants to M.B.

Correspondence and requests for material should be addressed to M.B. (e-mail: belfort@wadsworth.org).